Aquaporins and the vacuolar H^+-pyrophosphatase in *Phytophthora nicotianae*

Corinna S. Paeper

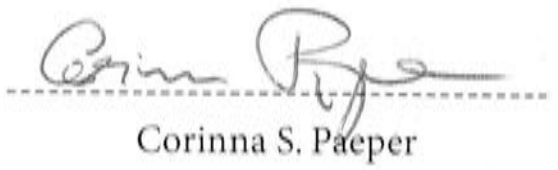
August 2006

THE AUSTRALIAN NATIONAL UNIVERSITY

A thesis submitted for the degree of Doctor of Philosophy of the Australian National University

Statement

**All research reported in this thesis is original and my own
except where acknowledgement is given,
and has not been submitted
for any other degree.**



Corinna S. Paeper

Acknowledgments

I would like to express my sincerest gratitude to my research supervisor Prof. Adrienne Hardham for her support and encouragement throughout the last 3.5 years. Without your enthusiasm, dedication and belief in my ability, this thesis would not have been completed.

Special thanks go to my advisors Dr Leila Blackman and Dr David Jones for guidance and advice when most needed.

Thanks also to the *Phytophthora* Molecular Genetics Network, and especially Prof. Brett Tyler, for providing financial support that allowed me to undertake a 2 week training visit at the VBI, Blacksburg, VA, USA in Prof. Brett Tyler's lab.

Many thanks also go to Val Mclean and Jan Elliott, our technical officers, for keeping the lab running, for preparation of northern blots and the generation of antibodies, and all the other 'little things'.

I would like to thank all other members of the lab over the years Dr Weixing Shan, Dr Dubravka Škalamera, Dr Reena Narayan, Dr Andrea Robold, Dr Miki Arikawa, Bettina Schulz, Patrick Golletz, Emma Landgren, Pamela Gan and Soma Chakrabarti, for advice, enlightening discussions and moral support.

Thanks to all other Plant Cell Biology group and RSBS people. Special thanks to Ginny Lim, Katherine Martin, Pinar Cinar and Apratim Chakrabarti for listening to my whinging and whining particularly in the final months of writing.

Special thanks to the countless people from the ANU Mountaineering and Scuba Club who made my outdoor adventures so special. I especially like to thank Annabel Battersby and Sam Keech-Marx for fun times on the MC executive, Paul Thomas and Dr Ernie Perkins for frequent coffee breaks, long chats in the final weeks and great diving trips, and Alison Shield and Chris Hepplewhite for letting me house-sit their fantastic dwelling.

Thanks also go to the 'Philosophy-Crew' for entertaining parties and weekly Fellow's and pizza get-togethers. Special thanks to Carl Brusse for sharing a good 2 years of his life with me, Ben Jeffares, Brett Calcott and Sam Lilie for talking to me, making me laugh, understanding me, listening to me, helping me with a lot of things, and mostly for 'just' being there for me!

Finally, I would like to thank my parents for their love, support and encouragement.

This thesis is dedicated to my parents

Abstract

The genus *Phytophthora*, class Oomycetes, comprises over 60 highly destructive species, which are able to infect vast numbers of agricultural, forest and horticultural plants and cause severe economic and ecological losses worldwide each year. The most notorious of all species, *Phytophthora infestans*, the causal agent of the late potato blight, caused widespread devastation in Ireland in the 1840s, and even today, 130 years after its discovery, remains a problem despite continuing efforts to develop effective control measures. Other economically and ecologically important species are, for example, *Phytophthora cinnamomi,* which has a host range of over 2000 plants and causes severe damage to Australian native ecosystems, *Phytophthora sojae*, the cause of soybean root and stem rot, and *Phytophthora nicotianae*, which infects up to 300 plants including a number of important citrus and solanaceous crops. Although destruction by some *Phytophthora* species can largely be contained with the few chemical fungicides that are available, resistant isolates have emerged over the years, especially in *P. infestans*. Therefore, the identification of potential drug targets that allow for control of these devastating pathogens is of the utmost importance.

The research presented in this study describes the characterisation and partial functional analysis of two gene families, the vacuolar H^+-translocating pyrophosphatases (V-PPase) and aquaporins, in the organism *P. nicotianae*. Studies focused on motile zoospores, the primary means of infection of susceptible host plants, as the genes were considered to possibly play a role in water expulsion vacuole (WEV) function of *Phytophthora* zoospores. Identification and characterisation of genes involved in osmotic homeostasis of zoospores would make a significant contribution towards the understanding of water expulsion through contractile vacuoles in general and would furthermore present a potential drug target for the control of *Phytophthora* species.

Chapter 3 describes the isolation and characterisation of the *P. nicotianae* vacuolar H^+-translocating pyrophosphatase gene (*PnVPP*). *PnVPP* was identified during the differential screening of randomly selected clones of a *P. nicotianae* λ-ZAP zoospore library and showed high homology to the AVP2 protein from *Arabidopsis thaliana*. Southern blot analysis revealed that *PnVPP* is a single copy gene in *P. nicotianae*.

Expression of the *PnVPP* transcript, as determined by RNA blot analysis of four asexual developmental stages of *P. nicotianae*, was up-regulated in zoospores and germinated cysts, moderate in vegetative hyphae and not detectable in sporulating hyphae. The complete cDNA sequence was obtained by 5' RACE and a full-length genomic clone was isolated and sequenced. Multiple sequence alignments with V-PPase proteins from other organisms and identification of highly conserved motifs within the amino acid sequence revealed that the PnVPP protein is likely to belong to the Type II group of V-PPases. Phosphatase activity in microsomal fractions of *P. nicotianae* zoospores was detected by a phosphatase activity assay, and the V-PPase type-specific inhibitor aminomethylenebisphosphonate (AMBP) was shown to effectively inhibit V-PPase activity in these fractions. *In vivo* inhibition experiments failed to show an effect of AMBP on zoospores. Monoclonal and polyclonal antibodies directed towards a highly conserved region of V-PPases were used to localise the protein in *P. nicotianae* zoospores. Immunofluorescence assays showed labelling of the plasma membrane and WEVs.

In Chapter 4, the isolation and characterisation of five aquaporin genes in *P. nicotianae (PnAq)* is described. Gene *PnAq2* was isolated from a *P. nicotianae* germinated cysts cDNA library and showed high homology to a *Gallus gallus* 'aquaporin 7-like protein'. Southern blot analysis showed the presence of at least four gene copies. Isolation of a *P. nicotianae* genomic BAC clone and the assembly of a 9 kb contig following subcloning and sequencing of fragments representing labelled bands in Southern analysis, revealed four aquaporin genes within the contig. A fifth gene was identified in a genomic subclone with unknown genomic organisation. Expression of the *PnAq* genes in *P. nicotianae* during the four asexual life cycle stages was analysed by northern blot and refined by a modified single-strand conformation polymorphism approach. Transcripts of all genes but *PnAq5* were observed at different stages throughout the life cycle. Polyclonal antibodies raised against a synthetic peptide comprising the consensus N- and C-terminus of the PnAq proteins showed an identical labelling pattern to the V-PPase antibodies.

Chapter 5 describes the mining of V-PPase and aquaporin genes in *Phytophthora* genome and EST databases, as well as the phylogenetic analysis of sequence data obtained. It was shown that it is likely that all *Phytophthora* genomes contain both a Type

I and a Type II V-PPase gene. In phylogenetic analysis, orthologues were clustered together and grouped with Type I and Type II proteins of other organisms. Database analysis showed that aquaporin genes are present in all *Phytophthora* species analysed. Surprisingly, genomes of *P. sojae* and *P. ramorum* showed over 20 aquaporin gene copies. In total, 51 *Phytophthora* aquaporin sequences were retrieved and clustered into five subgroups. Southern blot analysis with group-specific probes derived from *P. sojae* aquaporin sequences revealed that the *P. nicotianae* genome also contains genes belonging to the other groups. Phylogenetic analysis of *Phytophthora* aquaporins in the context of a large number of aquaporin sequences from other organisms showed a close relationship to a number of aquaglyceroporins indicating that *P. nicotianae* aquaporins might facilitate glycerol and/or water transport.

Further characterisation of these genes is required to fully understand their role in *Phytophthora* and will ideally comprise transformation and GFP-tagging experiments. The results of this study provide a solid framework for future analysis, and may help in the development of effective control measures for *Phytophthora* species.

Abbreviations

°C	Degree Celsius
µl	Microliter
µg	Microgram
2D	Two-dimensional
3D	Three-dimensional
5' RACE	Rapid amplification of cDNA ends
A_{260}	Optical density (absorbance) at 260 nm
aa	Amino acid
AMBP	Aminomethylenebisphosphonate
AP	Alkaline phosphatase
ATCC	American Type Culture Collection
BAC	Bacterial artificial chromosome
BLAST	Basic local alignment search tool
bp	Base pairs
BSA	Bovine serum albumin
CAP	Calf alkaline phosphatase
CDD	Conserved domain database
cDNA	Complementary DNA
cpm	counts per minute
C-terminus	Carboxy-terminus
CV	Contractile vacuole
DAPI	4',6-diamidino-2-phenylindole
DCCD	*N,N'*-dicyclohexylcarbodiimide
DEPC	Diethyl pyrocarbonate
dH_2O, ddH_2O	Distilled water, double-distilled water
DIC	Differential interference contrast
DIG	Digoxigenin
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleic triphosphate
DOE	Department of Energy
dsDNA	Double-stranded DNA
dsRNA	Double-stranded RNA
DTT	Dithiothreitol
EDTA	Ethylenediamine tetra-acetic acid
EGTA	Ethyleneglycol-bis(β-aminoethylether) tetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
EST	Expressed sequence tag
EtBr	Ethidium bromide
FITC	Fluorescein isothiocyanate
g	Gram
g	Gravity
GC	Germinated cysts
GFP	Green fluorescent protein

GSP	Gene specific primer
GUS	β-glucuronidase
h	Hour(s)
HEPES	N-(2-hydroxyethyl) piperazine-N'-(2-ethanesulphonic acid)
HRPO	Horse radish peroxidase
IPTG	Isopropyl-β-D-thiogalactoside
JGI	Joint Genome Institute
Kan^R	Kanamycin resistance gene
kb	kilobase
kDa	Kilodalton
KLH	Keyhole limpet hemacyanin
L	Litre
LB	Luria-Bertani broth
M	Molar
mA	Milliampere
Mbp	Mega basepairs
MBS	Maleimidobenzoyl-N-hydroxysuccimide ester
MCS	Multiple cloning site
MES	2-(N-morpholino)ethanesulphonic acid
MESG	2-amino-6-mercapto-7-methylpurine riboside
mg	Milligram
min	Minute(s)
MIP	Major intrinsic protein
mL	Millilitre
mM	Millimolar
MOPS	3-(N-morpholino) propanesulphonic acid
mRNA	Messenger-RNA
mV	Millivolt
MW	Molecular weight
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
NEM	*N*-ethylmaleimide
ng	Nanogram
NIP	NOD26-like intrinsic protein
nm	Nanometre
N-terminus	Amino-terminus
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
pBS	pBluescript plasmid
PBS	Phosphate buffered saline
PBSBG	PBS with BSA and gelatine
PBST	PBS with Tween20
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PFGD	*Phytophthora* functional genomics database
pfu	Plaque forming unit
PGC	*Phytophthora* Genome Consortium
pI	Isoelectric point

PIP	Plasma membrane intrinsic protein
PIPES	Piperazine-1,4-bis(2-ethanesulphonic acid)
pmol	Picomolar
PNP	Purine nucleoside phosphorylase
PP_i	Inorganic pyrophosphate
RNA	Ribonucleic acid
RNAi	RNA interference
RNase	Ribonuclease
rpm	Revolutions per minute
rRNA	Ribosomal RNA
RT	Reverse transcriptase, room temperature
RT-PCR	Reverse transcriptase-PCR
s	Second(s)
SAM	Sheep F(ab')2 anti-mouse antibody
SARa	Sheep F(ab')2 anti-rabbit antibody
SDS	Sodium dodecyl sulphate
SH	Sporulating hyphae
SIP	Small basic intrinsic protein
sPPase	Soluble inorganic pyrophosphatase
ss	Single-stranded
SSCP	Single-strand conformation polymorphism
TAE	Tris-acetate/EDTA
TE	Tris/EDTA
TBS	Tris buffered saline
TBST	Tris buffered saline with Tween 20
TEMED	Tetramethyl ethylene diamine
TIP	Tonoplast intrinsic protein
TMD	Transmembrane domain
TSS	Transcription start site
UTR	Untranslated region
UV	Ultra violet
V	Volt
V-ATPase	Vacuolar adenosine triphosphatase
VBI	Virginia Bioinformatics Institute
VH	Vegetative hyphae
VMD	VBI Microbial Database
V-PPase	Vacuolar H^+-pyrophosphatase
vs.	Versus
Vsv	Ventral surface vesicle
W	Watt
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
Z	Zoospore

Table of Contents

Statement ii

Acknowledgments iii

Abstract v

Abbreviations viii

Table of Contents xi

Chapter 1: Literature review

1.1 Phylogenetic classification of the genus *Phytophthora* 1

1.2 The genus *Phytophthora* 6

1.2.1 *Phytophthora nicotianae* 8

1.2.2 The life cycle of *Phytophthora* 10

1.2.2.1 Structural characterisation of Phytophthora *zoospores* 12

1.2.2.2 The WEV 13

1.3 Gene discovery and characterisation of *Phytophthora* genes 14

1.3.1 Genome organisation and functional genomics in *Phytophthora* 15

1.3.1.1 Functional analyses: Transformation, reporter genes and gene silencing 16

1.3.2 Characterisation of *Phytophthora* genes 23

1.4 Objectives of this study and experimental approach 24

Chapter 2: Materials and Methods

2.1 Materials 25

2.1.1 Enzymes, chemicals and molecular biology kits 25

2.1.2 Markers 25

2.1.3 Disposable materials 25

2.1.4 Antibiotics 26

2.1.5 Bacterial strains 26

2.1.6 Antibodies 26

2.2 Methods - Molecular Biology: Preparative DNA techniques 26

2.2.1 Preparation of plasmid DNA 26

2.2.2 Preparation of genomic DNA 27

2.2.3 Concentration of DNA 28

2.2.4 Phenol/chloroform extraction 28

2.2.5 DNA extraction from agarose gels 29

2.2.5.1 QIAquick Gel Extraction Kit 29

2.2.5.2 UltraClean™15 DNA Purification Kit 29

2.2.6 DNA probe synthesis 30

2.2.6.1 Megaprime DNA labelling 30

2.2.6.2 DIG-High Prime DNA Labeling 31

2.2.7 cDNA library screening 31

2.2.7.1 Cell preparation 31

2.2.7.2 First round of screening 31

2.2.7.3 Plaque purification 32

2.2.7.4 Phagemid excision 32

2.2.8 Isolation of genomic clones from a BAC library 33

2.2.8.1 Screening 33

2.2.8.2 Isolation of genomic clones 33

2.3 Methods – Molecular Biology: Analytical DNA techniques 34

2.3.1 Agarose gel electrophoresis 34

2.3.2 Visualisation and imaging of nucleic acids 34

2.3.3 Quantification of nucleic acids 34
2.3.4 PCR 35
2.3.5 Southern blot analysis 36
2.3.5.1 The transfer process 36
2.3.5.2 Pre-hybridisation and hybridisation 37
2.3.5.3 Image acquisition 37
2.3.5.3.1 ^{32}P-labelled probes 37
2.3.5.3.2 DIG-labelled probes 37
2.3.6 Restriction endonuclease digestion of DNA 38
2.3.7 Cloning 38
2.3.7.1 Converting 3'- and 5'-overhangs to blunt ends 38
2.3.7.2 Dephosphorylation of linearised plasmid DNA 38
2.3.7.3 Ligation 39
2.4 Methods – Molecular biology: RNA techniques 39
2.4.1 Total RNA isolation using TRIzol® reagent 39
2.4.2 cDNA synthesis 40
2.4.3 5' RACE 41
2.4.4 SSCP analysis 43
2.4.5 RNA separation by formaldehyde agarose gel electrophoresis 44
2.4.6 RNA Blot analysis 44
2.5 Methods - Protein techniques 45
2.5.1 Protein isolation 45
2.5.2 Peptide design for production of antibodies 45
2.5.3 Peptide conjugation 45
2.5.3.1 KLH coupling - MBS method 45
2.5.3.2 BSA coupling – Glutaraldehyde conjugation 46
2.5.4 Production of antibodies 46
2.5.4.1 Polyclonal antibodies 47
2.5.4.2 Monoclonal antibodies 47
2.5.5 Purification of polyclonal antibodies 47
2.5.6 Bradford assay 48
2.5.7 ELISA 49
2.5.8 Immunofluorescence microscopy 49
2.5.9 Immunodot blot 50
2.5.10 SDS-PAGE and western immunoblot 51
2.5.11 Peptide competition assay 51
2.6 Methods - *Phytophthora* and Bacteria 52
2.6.1 *Phytophthora* strain and culture conditions 52
2.6.1.1 Vegetative hyphae 52
2.6.1.2 Sporulating hyphae 52
2.6.1.3 Zoospores 53
2.6.1.4 Germinated cysts 53
2.6.2 Microsomal preparation of *Phytophthora* zoospores 53
2.6.2.1 Determination of pyrophosphatase activity 53
2.6.3 Video analysis of *Phytophthora* zoospores 55
2.6.4 Competent bacteria and transformation 56
2.6.5 Storage of *E. coli* cultures 56
2.7 Bioinformatic analysis 56
2.7.1 Design of custom primers and DNA sequencing 56
2.7.2 Sequence data analysis 57
2.7.3 Database analysis 57
2.7.4 Phylogenetic analysis 58

Chapter 3: Vacuolar H^+-translocating pyrophosphatase in *P. nicotianae*

3.1 Introduction 59

3.1.1 Transport across cell membranes 59
3.1.2 Pyrophosphatases 61
3.1.3 Cloning of V-PPases and their distribution in organisms 63
3.1.4 Topography of the V-PPase protein and conserved regions in the primary structure 68
3.1.4.1 Functional residues 72
3.1.4.2 V-PPase and its substrate – the role of Mg^{2+} 73
3.1.5 Enzyme regulation and inhibition of V-PPase activity 73
3.1.6 Potential role of V-PPase in osmoregulation 75
3.1.7 Aims of this study 76

3.2 Results 77

3.2.1 Analysis of cDNA clone DS311 77
3.2.2 RNA Blot Expression Analysis of cDNA clone DS311 79
3.2.3 In search of the 5' end: Screening of a zoospore λ-library and 5' RACE 80
3.2.4 Isolation, subcloning and sequencing of a *PnVPP* genomic clone 83
3.2.5 PnVPP protein, homologies and conserved domains 89
3.2.6 Pyrophosphatase activity assay on microsomal preparations of *P. nicotianae* zoospores and the influence of inhibitor substance AMBP 92
3.2.7 Overexpression of a PnVPP fusion protein for the generation of antibodies 96
3.2.8 Peptide design for the production of antibodies 98
3.2.9 Production and purification of polyclonal antibodies against two PnVPP peptides 99
3.2.10 Monoclonal antibody production against PnVPP 102
3.2.11 Immunolocalisation of the V-PPase in *P. nicotianae* zoospores 103
3.2.12 Western immunoblot analysis and peptide competition assays of polyclonal and monoclonal antibodies 105
3.2.13 Construction of a dsRNA construct for transformation assays in *P. nicotianae* 108

3.3 Discussion 111

3.3.1 Cloning, structural organisation and expression of the *PnVPP* gene in *P. nicotianae* 111
3.3.2 Characteristics of the PnVPP protein – conserved domains and its structure 113
3.3.3 Immunolocalisation of PnVPP 117
3.3.4 Is PnVPP involved in osmoregulation? 121
3.3.5 Gene silencing 123

Chapter 4: Aquaporins in *P. nicotianae*

4.1 Introduction 125

4.1.1 Aquaporins – An overview 125
4.1.2 Aquaporin structure – the hourglass model 127
4.1.3 Aquaporins in mammals 134
4.1.4 Aquaporins in plants 137
4.1.5 Aquaporins in microorganisms 139
4.1.5.1 Bacteria 140
4.1.5.2 Fungi 140
4.1.5.3 Algae and protozoa 142
4.1.6 Aims of this study 144

4.2 Results 145

4.2.1 Sequencing and alignment of cDNA clone WS86 145
4.2.2 Isolation, subcloning and sequencing of a *PnAq* genomic clone 147
4.2.3 PnAq proteins, homologies and conserved domains 153
4.2.4 Northern blot analysis of cDNA clone WS86 159
4.2.5 *PnAq* expression study using a modified SSCP approach 160
4.2.6 Peptide design for production of antibodies 166
4.2.7 Antibodies against two PnAq peptides 168

4.2.8 Western immunoblot analysis and peptide competition assays of polyclonal antibodies..... 171
4.2.9 Construction of a dsRNA construct for transformation experiments..... 173
4.3 Discussion 175
4.3.1 Cloning and structural organisation of the *PnAq* genes..... 175
4.3.2 Expression of the *PnAq* genes in *P. nicotianae* 177
4.3.3 Characteristics of the PnAq proteins – conserved domains and their structure..... 179
4.3.4 Immunolocalisation of PnAq 181
4.3.5 Are aquaporins involved in *P. nicotianae* osmoregulation? 183
4.3.6 Gene silencing 185

Chapter 5: Phylogenetic analysis of V-PPase and aquaporins in *Phytophthora* species
5.1 Introduction..... 187
5.1.1 Databases..... 187
5.1.1.1 The Phytophthora Genome Consortium (PGC)..... 187
5.1.2 Phylogeny..... 191
5.1.2.1 Phylogeny and the reconstruction of phylogenetic trees..... 191
5.2 Results 193
5.2.1 V-PPase..... 193
5.2.1.1 Database analysis..... 193
5.2.1.2 Comparison of proteins: Similarities and multiple sequence alignments..... 194
5.2.1.3 Phylogenetic analysis..... 198
5.2.2 Aquaporins..... 200
5.2.2.1 Database analysis..... 200
5.2.2.2 Protein characteristics..... 203
5.2.2.3 Multiple sequence alignments and conserved domains..... 205
5.2.2.4 Comparison of proteins: Identities and similarities..... 208
5.2.2.5 Southern blot analysis..... 210
5.2.2.6 Phylogenetic analysis..... 213
5.3 Discussion 214
5.3.1 V-PPase..... 214
5.3.2 Aquaporins..... 215
5.3.2.1 DB analysis..... 215
5.3.2.2 Protein characteristics and conserved domains..... 217
5.3.2.3 Southern blot analysis..... 219
5.3.2.4 Phylogenetic analysis..... 219

Chapter 6: General discussion and conclusions
6.1 V-PPases, aquaporins and the WEV 221
6.2 Synteny of aquaporin genes between *Phytophthora* species 223
6.3 Future directions..... 224

Bibliography 226
Appendix A Buffers and solutions 251
Appendix B Bacterial strains and media 260
Appendix C Antibodies 262
Appendix D *Phytophthora* culture media 263
Appendix E Primers 264
Appendix F V-PPase sequence details 267
Appendix G Aquaporin sequence details 279
Appendix H Database and phylogeny manual 311

Chapter 1

Literature review

Chapter 1: Literature review

The genus *Phytophthora* includes many economically and ecologically important pathogens. In the 1840s, the plant pathogen *Phytophthora infestans*, the causal agent of late blight of potato, caused widespread devastation in Ireland by destroying the entire potato crop of two consecutive years. About one million people died of starvation and approximately the same number of people emigrated to the rest of Europe and the USA as a result of the following famine (Erwin and Ribeiro, 1996). Within 20 years, Ireland's population was diminished from 8.2 million to 5.8 million people (Duncan, 1999). No other known plant pathogen has had such an impact on human life to date.

Even though about 130 years have passed since *P. infestans* was linked to the devastating disease by Anton de Bary in 1876 (reviewed in Erwin and Ribeiro, 1996), late blight epidemics still occur all over the world. Despite intensive genetics and breeding research, potato cultivars with persistent late blight resistance have not been developed (Irwin *et al.*, 1995), and finding effective control measures for this pathogen is an ongoing challenge. Even today, costs for losses of potato crop due to potato blight and the effort of controlling it, amount to approximately five billion US dollars per year worldwide (Birch and Whisson, 2001). However, *P. infestans* is not the only *Phytophthora* species that causes significant damage to plant species worldwide.

1.1 Phylogenetic classification of the genus *Phytophthora*

Members of the genus *Phytophthora* are grouped within the algae-like class of Oomycetes ('egg fungi'), which belong to the kingdom Stramenopila. Stramenopiles are eukaryotic organisms which are generally characterised as having flagella with tripartite, tubular hairs (also called mastigonemes), and the kingdom not only includes Oomycetes but also heterokont algae, ranging from the enormous brown algae (kelp) to the unicellular diatoms (Cavalier-Smith, 1993; Sogin and Silberman, 1998; Patterson, 1999).

The class Oomycetes is divided into three orders, Peronosporales, Rhipidiales, and Saprolegniales, which represent approximately 65 genera and around 700 species (Table 1.1, Guarro *et al.*, 1999). Many of its members are saprophytic and pathogenic microorganism, which affect a wide range of plants, animals and aquatic organisms, causing economic and ecological losses worldwide. Species of the order Saprolegniales and Rhipidiales (e.g. *Saprolegnia*, *Achlya*, *Rhipidium*) are mainly aquatic (also called

'water molds') and parasitic to a number of freshwater fish, whereas Peronosporales species are plant pathogens (downy mildew, white rust, root rot). The genus *Phytophthora* is placed in the order Peronosporales, which contains predominantly parasitic 'lower fungi' of the families Albuginaceae, Peronosporaceae and Pythiaceae.

The taxonomy of the Oomycetes is in a constant state of flux. Historically, oomycetes which resemble true fungi in appearance and share similar nutritional modes (Hardham *et al.*, 1994) were classified as lower fungi and placed in the fungal kingdom. However, there are a number of morphological and biochemical differences between the true fungi and oomycetes as outlined in Figure 1.1. For example, the cell wall of oomycetes is not composed of chitin, but instead contains β-1,3-glucan, β-1,6-glucan and cellulose (Bartnicki-Garcia and Wang, 1983), and mycolaminarin, a β-1,3-glucan, is the characteristic carbohydrate storage molecule (Erwin and Ribeiro, 1996). Furthermore, oomycetes have mitochondria with tubular cisternae (as opposed to plate-like), have a diploid life cycle and non-septated hyphae. *Phytophthora* and *Pythium* species in addition are unable to synthesise sterols, and thus require an exogenous source of β-hydroxy sterols for sporulation (cited in Erwin and Ribeiro, 1996).

Table 1.1 Taxonomy of *Phytophthora* (Cooke *et al.*, 2000; Riethmuller *et al.*, 2002; Kroon *et al.*, 2004)

Kingdom	Phylum	Class	Order	Family	Genus
Stramenopiles	Bacillariophyta				
	Chrysophyta				
	Phaeophyta				
	Oomycota	**Oomycetes**	Rhipidiales		
			Saprolegniales		
			Peronosporales	Albuginaceae	
				Pythiaceae	
				Peronosporaceae	***Phytophthora***
					Bremia
					Plasmopara
					Peronospora
					Perenophythora
					Sklerosporales

Feature	Oomycete	True fungi
Neighbouring taxonomic groups	Diatoms and golden-brown algae	Animals
Hyphal architecture	Aseptate and coenocytic tubular hyphae	Either single cell or septated hyphae, with one or more nuclei per compartment
Ploidy of vegetative hyphae	Diploid, except for transient haploid nuclei in gametangia	Typically haploid or dikaryotic; often with a stable or semi-stable diploid stage following mating
Typical size of genome	50–250 Mb	10–40 Mb
Major glucans in cell walls	Cellulose (β-1,4-linked glucose), and β-1,3, and β-1,6-linked glucose polymers	Usually chitin (β-1,4-linked *N*-acetylglucosamine) and/or chitosan (β-1,4-linked glucosamine), often with other β-1,3, and β-1,6 glucans
Pigmentation	Usually unpigmented	Very common in hyphae or spores, or secreted (for example, melanin, carotenoids and others)
Toxic secondary metabolites	None described	Common (typically aromatic, heterocyclic compounds)
Mating hormones	Non-peptide, probably lipid-like	Usually small peptides or lipopeptides
Predominant asexual spore	Undesiccated, unicellular sporangia (multinucleate cells)	Desiccated single or multicellular conidia (one nucleus per cell)
Motile asexual spores	Nearly universal, biflagellated zoospore	Uncommon, only in chytrids, which are monoflagellate
Sexual spores	Oospores, formed on the termini of specialized hyphae, each containing one viable zygotic nucleus	Various types, often formed in large numbers within complex enclosures (for example, perithecia, mushroom caps and others)
Major energy reserves used by spores	Mycolaminarin and lipid, possibly polyphosphate	Glycogen and trehalose, also sugar alcohols and lipid

Figure 1.1 Major differences between oomycete 'fungi' and true fungi. (Source: Judelson and Blanco, 2005)

More recent phylogenetic studies of mitochondrial, nuclear and ribosomal DNA (Van der Auwera and De Wachter, 1996; Cooke *et al.*, 2000; Riethmuller *et al.*, 2002; Kroon *et al.*, 2004) confirmed the divergence between Oomycetes and true fungi (Figure 1.2). Furthermore, the genus *Phytophthora*, previously located in the family Pythiaceae together with the genus *Pythium*, was moved to the family Peronosporaceae which also comprises the genera *Bremia*, *Plasmopara* and *Peronospora* (Figure 1.3).

The features displayed by the oomycetes and their phylogenetic distance from true fungi has to be taken into consideration when devising efficient control measures. Many fungicides commonly used for containment of true fungi dissemination fail to work on oomycetes. Azole fungicides, for example, which target the ergosterol biosynthesis pathway in fungi, are ineffective against oomycetes as they are devoid of this pathway (e.g. Erwin and Ribeiro, 1996). A number of systemic compounds (phosphorous acid and phenylamide fungicides) are the primary chemicals for the control of diseases caused by oomycete species (Irwin *et al.*, 1995). Metalaxyl, a phenylamide which inhibits ribosomal RNA (rRNA) synthesis, is the most widely used compound against these pathogens (Erwin and Ribeiro, 1996). However, resistance against metalaxyl has emerged in

numerous pathogens over the years and it is important to carefully monitor and manage the use of these chemicals to avoid complete resistance and retain their usefulness against Oomycete species (Irwin *et al.*, 1995; Erwin and Ribeiro, 1996).

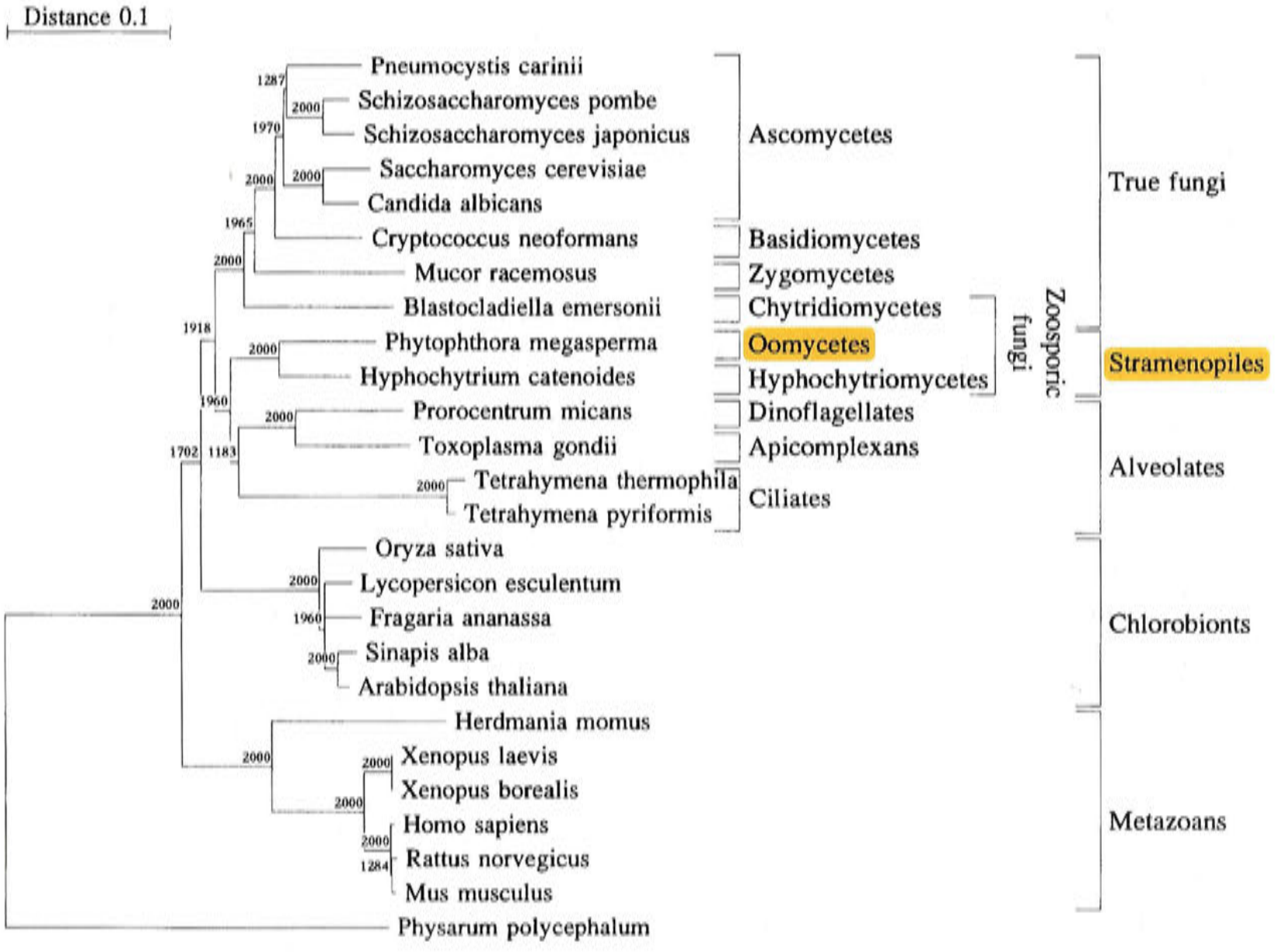


Figure 1.2 Neighbor-joining tree constructed on the basis of a combination of 5.8S and 18S rRNA sequences. Bootstrap analysis was performed on 2,000 samples, and values are shown at the internodes. (Source: Van der Auwera and De Wachter, 1996)

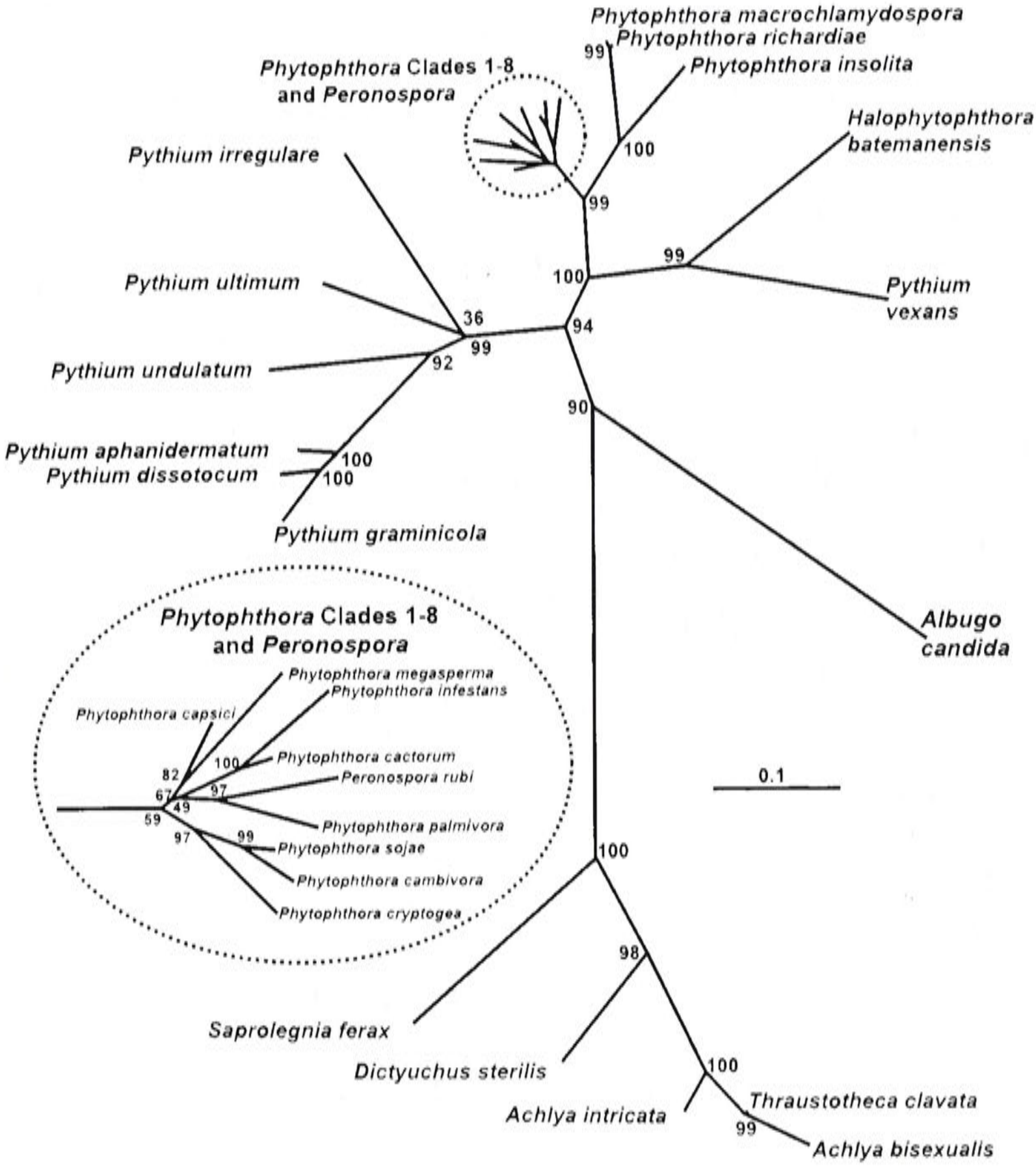


Figure 1.3 Phylogram of genera and species of the Saprolegniales and Peronosporales obtained by DNA distance-based analysis of the combined 5.8S subunit and ITS2 regions of the genomic rRNA tandem gene repeat. The numbers at the branch points indicate the percentages of bootstrap values (based on 500 bootstraps). The inset shows the details of the relationships among eight *Phytophthora* species and a *Peronospora* species. (Source: Cooke *et al.*, 2000)

1.2 The genus *Phytophthora*

The genus *Phytophthora* contains more than 60 species, all of which are destructive pathogens of agriculture, horticulture or forestry plants (Table 1.2). The name is derived from the greek and literally translates into 'plant destroyer' (*phyto* – plant, *phthora* – destroyer) (Erwin and Ribeiro, 1996). Members of the genus cause severe economic and ecological losses to agriculture and forest industries in Australia and worldwide (Irwin *et al.*, 1995; Erwin and Ribeiro, 1996).

Table 1.2 Economically important *Phytophthora* species

Phytophthora species	Major host plant(s)
Phytophthora infestans	Potato and tomato
Phytophthora cinnamomi	Avocado, pineapple, avocado, numerous Australian native forest species
Phytophthora sojae	Soybean
Phytophthora ramorum	Oak
Phytophthora nicotianae	Tobacco, tomato, citrus species
Phytophthora capsici	Pepper and Cucurbitaceae species
Phytophthora palmivora	Cocoa
Phytophthora megasperma	Brassica species
Phytophthora fragariae	Strawberry

The host range of *Phytophthora* species varies significantly. *P. infestans*, the most notorious species of the whole genus, only affects potato and tomato crops, whereas another species, *P. cinnamomi*, has an extremely wide host range with an estimated 2000 susceptible plants (Hardham, 2005). *P. cinnamomi*, first recorded in 1922 in Sumatra, causes severe damage to forests in Australia, Europe and the USA (Irwin *et al.*, 1995; Erwin and Ribeiro, 1996). Also known as 'dieback' or root rot, *P. cinnamomi* has a major impact on not only a wide range of economically important crop plants such as pineapple, chestnut, cherry, papaya, avocado in Australia, but also on native vegetation (e.g. jarrah, banksia, grass trees), destroying whole ecosystems especially in Western Australia, Victoria and Tasmania (Figure 1.4, Irwin *et al.*, 1995; Erwin and Ribeiro, 1996; Hardham, 2005).

Figure 1.4 Grass trees (*Xanthorrhoea australis*) in the Brisbane Ranges, Victoria. (a) Healthy forest versus (vs.) (b) *P. cinnamomi* infected area. (Photographs by A. Hardham)

Although *P. sojae* can infect a number of host plants, soybean remains the only agricultural and economically important susceptible crop plant. Root and stem rot caused by *P. sojae*, which can infect soybean plants at all growth stages, is widely distributed in soybean plantations in Canada and the USA, and has also become a problem in eastern parts of Australia (Erwin and Ribeiro, 1996). Sudden oak death, caused by *P. ramorum*, was first reported in California in 1995, and is now destroying the oak ecosystem of central coastal areas of California, and threatening to spread further inland and to the east coast (Knight, 2002; Rizzo *et al.*, 2005).

1.2.1 *Phytophthora nicotianae*

P. nicotianae was first isolated from *Nicotiana tabacum* from Sumatra and Java by Breda de Haan in 1896 (Erwin and Ribeiro, 1996). A range of synonyms for the species appeared over the years, and until 1963 *P. parasitica* (Dastur 1913) rather than *P. nicotianae* was the widely accepted name worldwide. Even though the species was re-described in 1963 by Waterhouse and the name *P. nicotianae* officially reinstated, *P. parasitica* is still used by some researchers predominantly in the USA some 40 years later (Hall, 1993; Erwin and Ribeiro, 1996).

P. nicotianae has a wide host range, possibly infecting up to 300 different plant species worldwide, however, it has been noted that some isolates are specifically virulent to a particular host (Erwin and Ribeiro, 1996). This destructive pathogen causes root rot in the majority of plants, nevertheless, leaves, seedlings or stems of some hosts are also affected. Attacking a number of important citrus and solanaceous crops (e.g. tobacco, tomato, capsicum - Table 1.3, Figure 1.5, Figure 1.6), *P. nicotianae* causes significant economical losses worldwide (Hall, 1993; Erwin and Ribeiro, 1996).

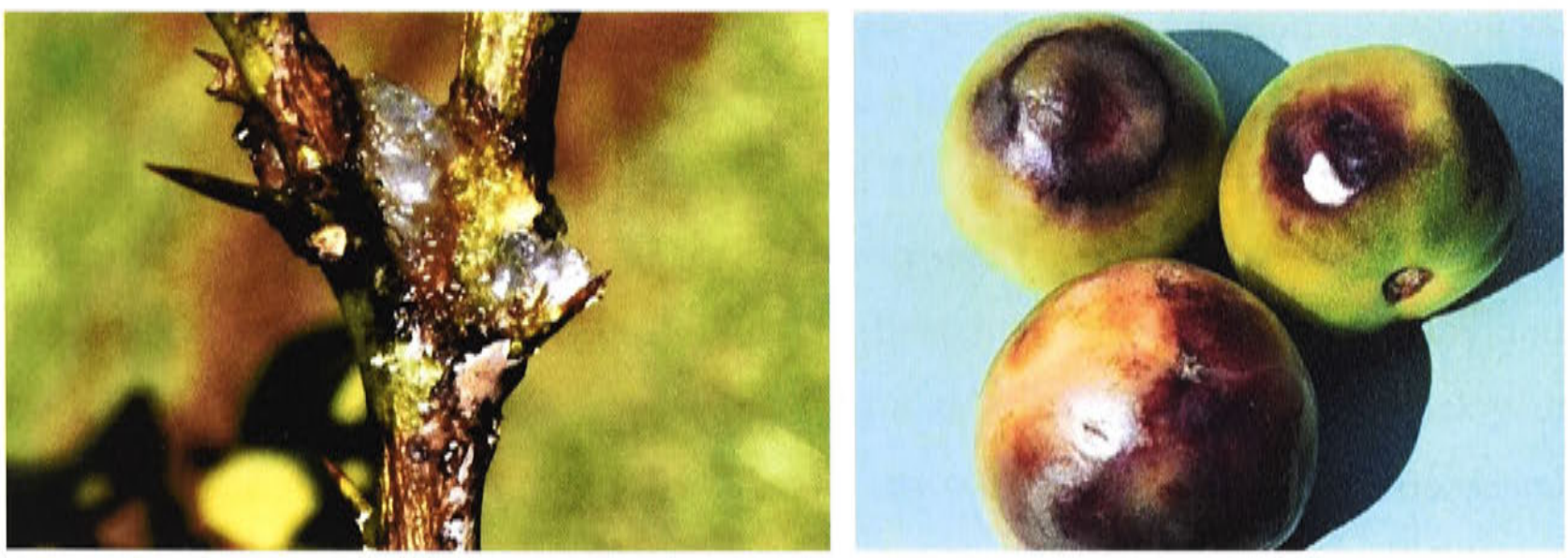
Table 1.3 Ecologically important crop plants susceptible to *P. nicotianae* and occurring symptoms upon infection

Plant	Symptoms
Citrus species	Root and fruit rot
Tobacco	Black shank disease (root rot, lesions, wilting)
Tomato	Buckeye and stem rot
Capsicum	Root and stem rot

In Australia, *P. nicotianae* was first described in *Primula* species (primrose) in 1927 by Brittlebank and Fish, and in English walnut in 1929 by Cookson (Erwin and Ribeiro, 1996). The majority of reports concerning *P. nicotianae* infections occurred in the 1950's, on citrus, tomato, castor bean and alfalfa (Erwin and Ribeiro, 1996). With regards to native vegetation, only a few *Banksia* species are hosts for *P. nicotianae* (Erwin and Ribeiro, 1996), however, in the light of the devastation caused by *P. cinnamomi* in Australian native fauna, the infestation by *P. nicotianae* is generally neglected. The impact of *P. nicotianae* in Australia is mainly of economical nature. It is a threat to citrus plantations in the sub-tropical regions of eastern Australia, and to tobacco farms in Queensland and Victoria (Industry Commission, 1994; Winstanley *et al.*, 1995). A survey carried out in 1993 reported annual yield losses of up to 15% in *Citrus* species and 6% in

tobacco production (Cahill, 1993). Furthermore, disease outbreaks in tomato plantations due to soil flooding by for example heavy rain can lead to losses up to 25% (cited in Washington and McGee, 2000).

Figure 1.5 Infection of *N. tabacum* with *P. nicotianae*. (Source: www.forestryimages.org, Clemson University - USDA Cooperative Extension Slide Series)



a) b)

Figure 1.6 *P. nicotianae* infections. (a) Gummosis on infected citrus stem. (Source: http://www.phytid.org/CPC/pics/p.nicotianae.htm), (b) buckeye rot of tomato fruit. (Source: http://ftsg.ifas.ufl.edu/BUCDBC.htm)

1.2.2 The life cycle of *Phytophthora*

Phytophthora species are able to reproduce sexually and asexually (Figure 1.7), however, the asexual pathway is the main cause for rapid and widespread infections. *Phytophthora* grows as vegetative mycelium in the soil or on plant tissues and forms sporangia under nutrient-deprived conditions. Dissemination and initiation of infection for most species is achieved by uni-nucleate, motile and wall-less zoospores, which are produced by sporangial cleavage upon environmental triggers such as drop in temperature or availability of free water. Released into the surrounding medium, the biflagellate zoospores can swim for hours or even days using endogenous food reserves (Carlile, 1983). Zoospores can swim actively towards the roots of potential host plants, as they respond to chemical gradients (Khew and Zentmyer, 1973; Morris and Ward, 1992; Cahill and Hardham, 1994) and electrical fields (Morris and Gow, 1993; van West *et al.*, 2002). At the root surface, the zoospores encyst: they detach the flagella, a cell wall is formed and adhesive material that sticks the cysts to the adjacent surface is secreted (Carlile, 1983; Hemmes, 1983). This process is completed within about 5-6 minutes (Gubler and Hardham, 1988). Approximately 20-30 minutes after encystment, the cysts germinate and the germlings penetrate the plant cell wall and colonise the plant, in the process acquiring the nutrients needed for growth and sporulation (Erwin and Ribeiro, 1996; Hardham and Hyde, 1997). Therewith, the cycle is completed and can start again.

In some *Phytophthora* species chlamydospores are produced in addition to zoospores. Chlamydospores are larger than zoospores and are thought to be the long-term survival propagules in the soil (also called resting spore). They are formed terminally at the tips of hyphae, or between the tip and the base of the hyphae (intercalary), and are separated from the mycelium by septa (Erwin and Ribeiro, 1996; McCarren *et al.*, 2005). Thin- and thick-wall chlamydospores have been observed and it was proposed that thick-wall spores can persist in soil for long times and can withstand extreme environmental conditions, with survival rates of up to 4-6 years (summarised in Erwin and Ribeiro, 1996). However, direct evidence of this phenomenon is sparse and further research is required to elucidate factors like long-term survival, and germination behaviour of chlamydospores (discussed in McCarren *et al.*, 2005).

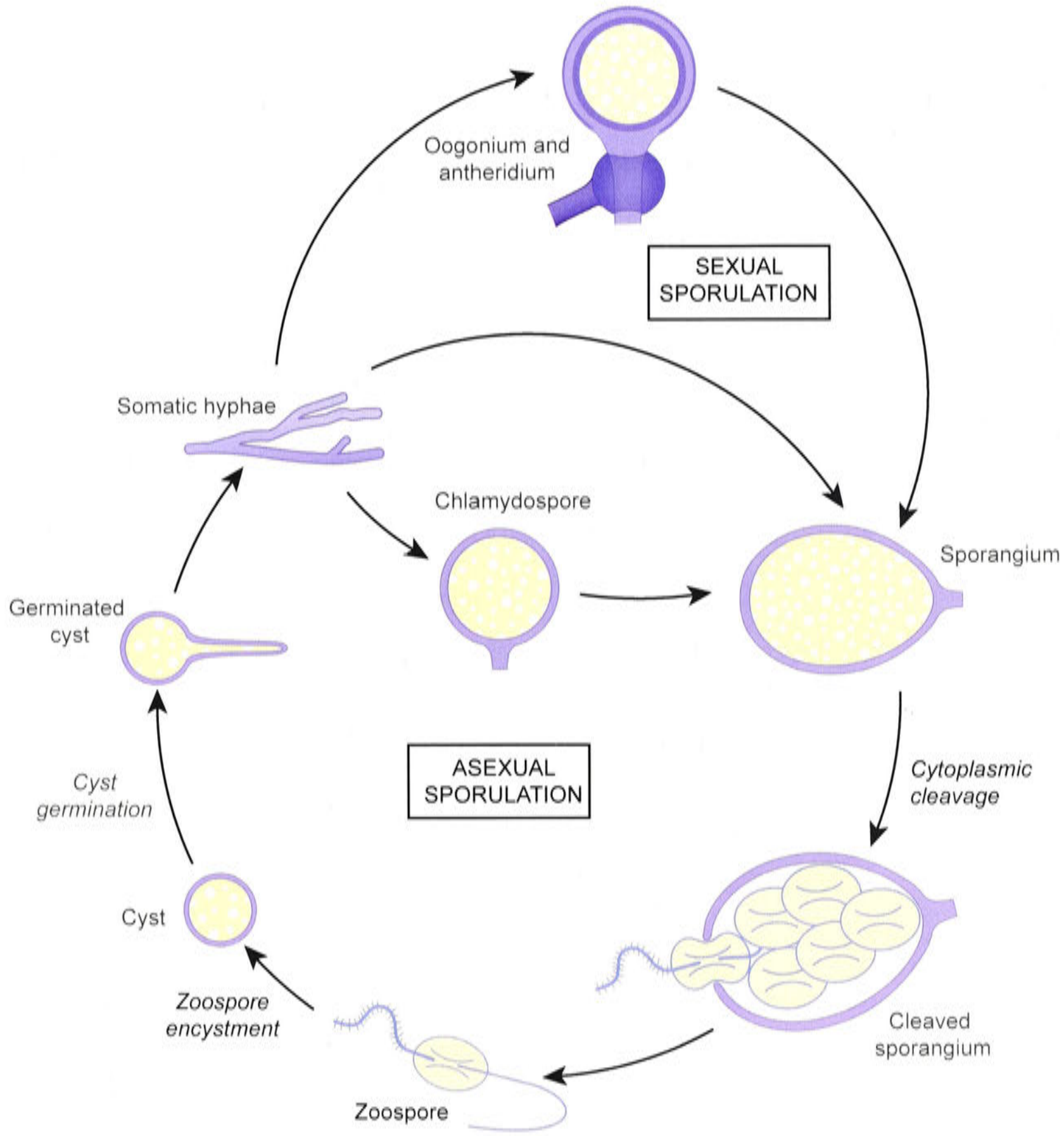


Figure 1.7 Sexual and asexual life cycle of *P. cinnamomi*. See text for details. (Graphic by A. Hardham)

1.2.2.1 Structural characterisation of Phytophthora *zoospores*

Phytophthora species produce wall-less zoospores (Figure 1.8) that are less than 10 µm in diameter and ovoid in shape with a longitudinal groove along what is termed the ventral surface (Carlile, 1983; Hemmes, 1983). The two flagella which emerge from basal bodies in the groove are responsible for zoospore motility. The posterior whiplash flagellum acts as a rudder and allows the zoospore to turn as it swims, whereas the anterior flagellum, which bears hair-like mastigonemes, generates the forward thrust (Carlile, 1983; Cahill *et al.*, 1996). Zoospores can swim through water-logged soil for hours along a helical path, and can cover several cm at a speed of 100-200 µm/sec depending on temperature and species (Duniway, 1976; Carlile, 1983; Cahill *et al.*, 1996).

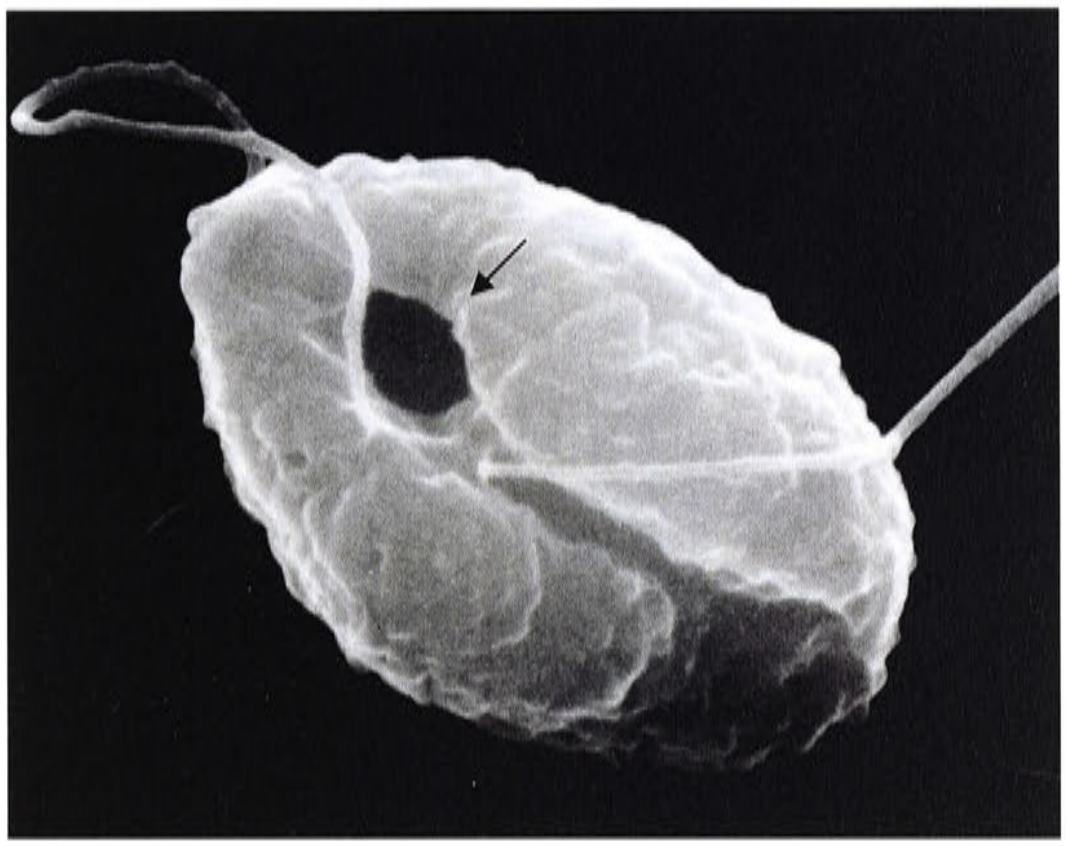


Figure 1.8 Scanning electron micrograph of a *P. cinnamomi* zoospore. The ventral surface of the zoospore is shown. The two flagella (anterior and posterior) emerge from a small ridge in the groove. The water expulsion vacuole (WEV, arrow) is thought to be involved in zoospore osmoregulation. (Source: Hardham, 1987)

Ultrastructural investigations have shown that the zoospore cytoplasm is highly structured (Hardham, 1987). At the apex of the pear-shaped nucleus, which is located in the middle of the zoospore, the basal bodies for the two flagella as well as Golgi bodies are found. In the anterior part of the cytoplasm a WEV, which is thought to be involved in the process of osmoregulation in zoospores, is found (Bouck, 1972; Hardham, 1987). Furthermore, there are mitochondria as well as three distinct types of peripheral vesicles in the remaining cytoplasm (Hemmes, 1983; Hardham, 1987). These vesicles show

different polarities in their distribution within the zoospore cortex. The large peripheral vesicles are elliptical in cross section and occur with their long axis at right angles to the cell surface. They are distributed all over the cell periphery except in the groove region (Hardham, 1987; Gubler and Hardham, 1988), and are thought to play a role in nutrient supply during early infection (Gubler and Hardham, 1990). The two types of small peripheral vesicles are described as dorsal and ventral vesicles as they occur mainly on the dorsal or ventral side of the zoospore, respectively. Contents of the dorsal vesicles are secreted upon encystment of the zoospore (Gubler *et al.*, 1989), and form a layer that may serve a protective function against desiccation. The contents of the ventral vesicles (Vsv) are also secreted during encystment of zoospores (Hardham and Gubler, 1990). The results of molecular and cell biology studies suggest, that the secreted Vsv contents constitute the adhesive material that glues the spore to the host's surface (Hardham and Gubler, 1990; Robold and Hardham, 2005).

1.2.2.2 The WEV

Due to the lack of a cell wall, *Phytophthora* zoospores face a major challenge in maintaining their cell volume and integrity under hypo-osmotic conditions found in soil water. Zoospores are subjected to a continuous influx of water due to a higher solute concentration inside the cell compared to the surrounding water. If cells are to maintain a steady internal solute concentration, water must be removed from the cell (against its concentration gradient) at the same rate at which it enters. It is thought that zoospores counteract the influx of water by sequestering and then pumping the water out of the cell via the WEV. The WEV in *Phytophthora* zoospores is crucial for their survival since without it, the zoospores would swell and lyse. The WEV repeatedly fills with water from the cytoplasm and expels its contents to the surrounding medium, completing a cycle approximately every 6 seconds (Mitchell and Hardham, 1999). Ultrastructural studies revealed that the WEV, located at the anterior end of the zoospore, consists of a central vacuole (bladder) which is surrounded by a reticulum of tubules and small vesicles (spongiome) (Figure 1.9, Patterson, 1980; Cho and Fuller, 1989). The process of water sequestration from the cytosol is thought to occur at the spongiome. In *Phytophthora* and other protists, the membranes of the spongiome have been shown to contain a vacuolar H^+-ATPase (V-ATPase) which may power water uptake (Heuser *et al.*, 1993; Fok *et al.*, 1995; Temesvari *et al.*, 1996; Mitchell and Hardham, 1999). The exact molecular mechanism for water sequestration is not known. However, it is possible that a

pH gradient and/or membrane potential generated by the V-ATPase drives the transport of osmolytes into the lumen of the spongiome (Nelson and Klionsky, 1996; Finbow and Harrison, 1997; Stevens and Forgac, 1997). This accumulation of osmolytes could then facilitate water uptake by the spongiome from the cytoplasm down a chemo-osmotic gradient.

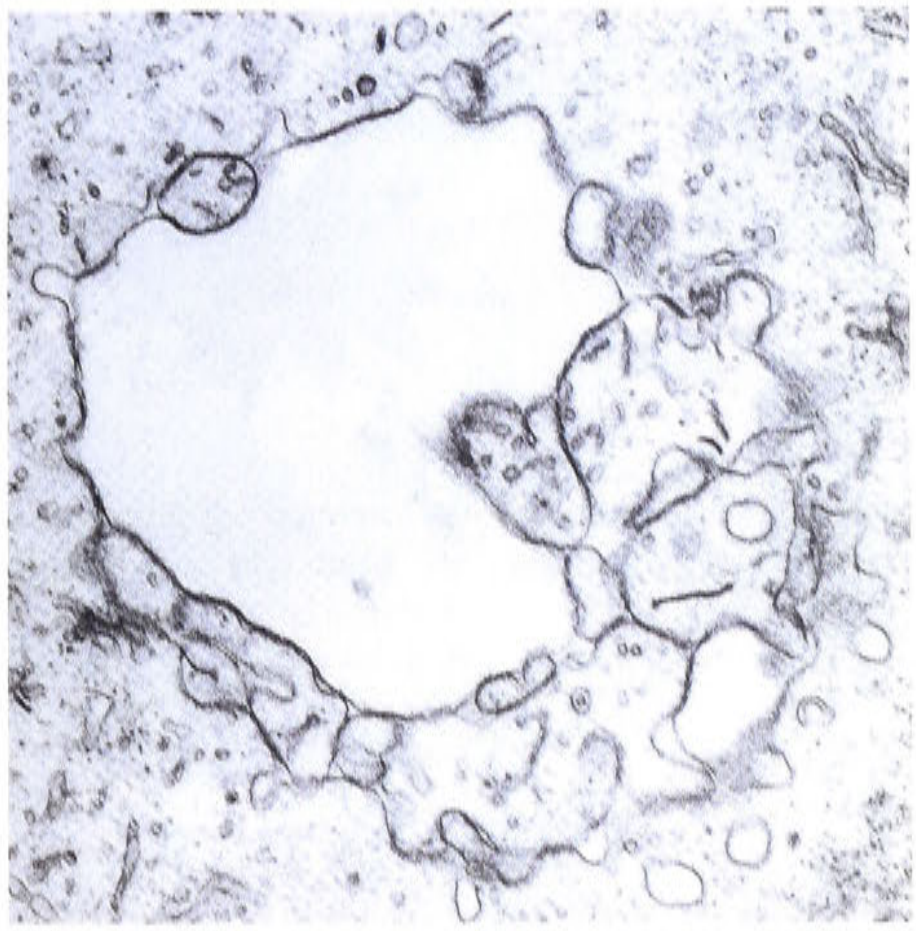


Figure 1.9 Transmission electron micrograph of a *Phytophthora* zoospore WEV. (Source: Mitchell and Hardham, 1999)

1.3 Gene discovery and characterisation of *Phytophthora* genes

Over the past few years, the availability of genomic data has grown exponentially. In combination with improved laboratory techniques, new possibilities for gene discoveries and their functional analysis have been created. This development will give new insight into the structure and organisation of pathogen genomes and will improve the understanding of the molecular basis of pathogenicity and host specificity of a number of *Phytophthora* species.

Kamoun *et al.* (2002) proposed a sequence-to-phenotype paradigm (Figure 1.10) for the discovery of novel virulence and avirulence genes in *Phytophthora* species, which links bioinformatic analysis approaches (i.e. database mining) with functional assays to characterise genes and determine their function and importance in the pathogen.

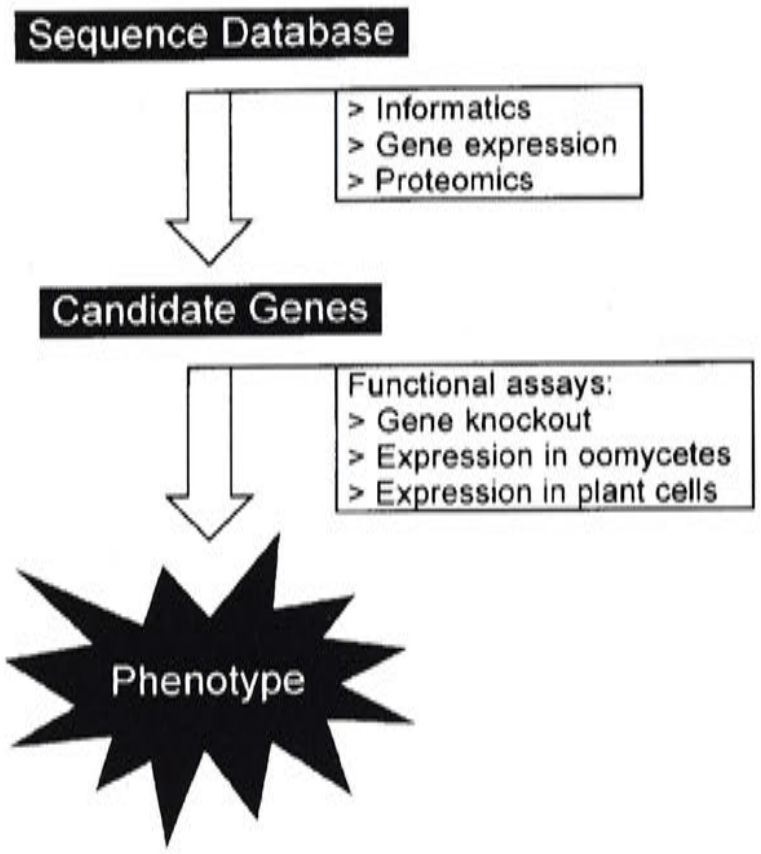


Figure 1.10 Flowchart illustrating the sequence-to-phenotype paradigm as applied to *Phytophthora*. The two-step process consists first of mining for candidate genes from sequence databases using informatics, gene expression proteomics, or a combination of these technologies. Then, functional assays are applied to the candidate genes to identify those that confer a desirable phenotype. (Source: Kamoun *et al.*, 2002)

1.3.1 Genome organisation and functional genomics in *Phytophthora*

Over the past two decades, increased awareness of Oomycetes and their impact on ecosystems and economy have led to Oomycete-specific molecular methodologies and advances in genomic analyses and functional genomics. Novel technologies such as Oomycete-specific transformation protocols, gene silencing and the use of reporter genes have only been available since the early 1990s. These methodologies in combination with the collection of a number of expressed sequence tags (ESTs) and genomic databases have opened up new resources for the molecular study of these pathogens and their host interactions.

EST clones from a variety of *Phytophthora* species, and developmental and infection stages have been generated over the past 10 years. Large EST data sets are now on hand in the form of compiled databases and include over 75,000 EST clones from *P. infestans* (Kamoun *et al.*, 1999; Randall *et al.*, 2005) and approximately 21,000 ESTs from *P. sojae* (Qutob *et al.*, 2000; B. M. Tyler, unpublished data; http://staff.vbi.vt.edu/estap/). More recently, smaller projects have emerged and supplied the research community with EST data of *P. nicotianae* (Shan *et al.*, 2004; Škalamera *et al.*, 2004; Panabieres *et al.*, 2005) and

P. cinnamomi (R. Narayan, J. Marshall and A. Hardham, unpublished results). Genomic BAC (bacterial artificial chromosome) libraries have also been generated for some *Phytophthora* species, including *P. infestans* (Randall and Judelson, 1999; Whisson *et al.*, 2001), *P. nicotianae* (Shan and Hardham, 2004), *P. sojae* and *P. ramorum* (http://pmgn.vbi.vt.edu/choose_link.php?id=resource_list). More recently, annotated sequences of the whole genomes of *P. sojae* and *P. ramorum* have become available on the web site of the Department of Energy (DOE), Joint Genome Institute (JGI) (http://genome.jgi-psf.org/).

Phytophthora species show a broad range of genome sizes and these are larger than those of fungi and other microorganisms (Jiang *et al.*, 2005). The genome of *P. infestans* was estimated to be 237 Mbp by 4',6-diamidino-2-phenylindole (DAPI) microfluorometry (Tooley and Therrien, 1987), and a large-scale EST project, using data from 19 different conditions of growth and development, predicted a unigene set of 18,256 (Randall *et al.*, 2005). The genome sequencing projects of *P. sojae* and *P. ramorum* revealed estimated genome sizes of 95 Mbp with a total of ~19,000 genes, and 65 Mbp with ~16,000 genes, respectively (Tripathy *et al.*, 2006, http://genome.jgi-psf.org/sojae1/sojae1.home.html, http://genome.jgi-psf.org/ramorum1/ramorum1.home.html). The 10-fold coverage of the *P. nicotianae* genome by a BAC library predicted a genome size similar to that of *P. sojae* of 95.5 Mbp (Shan and Hardham, 2004). *Phytophthora* genomes in general are characterised by an abundance of repetitive sequences variable in copy numbers, and it was estimated that repeats represent more than half of the *P. infestans* and *P. sojae* genomes (Judelson and Randall, 1998; Kamoun, 2003). Interestingly, even though *P. infestans* and *P. sojae* show vastly differing genome sizes, they contain approximately the same number of genes. It is thought that genes are generally located within high-density gene-islands that are separated by clusters of repetitive sequences (e.g. *PsojNIP* family of *P. sojae*, *ipiO* and *ipiB* of *P. infestans*, in Kamoun, 2003).

1.3.1.1 Functional analyses: Transformation, reporter genes and gene silencing

The availability of genomic data and the development of bioinformatic tools to effectively mine the databases, facilitated the discovery of innumerable 'genes of interest'. For the functional assessment of these genes, efficient assays at the ultrastructural and biochemical levels, such as localisation studies with reporter genes and gene/protein knock-out studies via gene silencing, are required. All these techniques rely heavily on a

DNA transformation system. Unfortunately, this step proved to be the bottleneck in *Phytophthora* research, as efficient and reliable transformation protocols are still not readily available.

As oomycetes are very different from true fungi, novel transformation protocols and vectors had to be developed. A range of vectors as well as several promoters and terminators, such as ham34 and hsp70 from the downy-mildew oomycete *Bremia lactucae*, were tested thoroughly in early transformation experiments (e.g. Judelson *et al.*, 1991). The standard transformation protocol for *P. infestans*, first described by Judelson *et al.* (1991), used DNA-liposome complexes in the presence of polyethylene glycol (PEG) and $CaCl_2$ to transform protoplasts. Drawbacks of the technique were, however, a notoriously low transformation efficiency, high rates of tandem integration into the genome and no homologous recombination events (Kamoun, 2003). Additionally, the key enzyme for the production of protoplasts, Novozym 234 (Novo Nordisk, Denmark), was withdrawn from the market in 2000 (Birch and Whisson, 2001; van West et al., 2003). Thus, the transformation protocol had to be amended and different methods were subsequently investigated. Alternative protocols include transformation with *Agrobacterium tumefaciens* (Vijn and Govers, 2003), microprojectile bombardment (Cvitanich and Judelson, 2003b) and electroporation of zoospores (B. M. Tyler, unpublished data; Latijnhouwers and Govers, 2003).

Transformation protocols have successfully been used to express reporter genes such as β-glucuronidase (GUS), β-galactosidase and green fluorescent protein (GFP) in a number of *Phytophthora* species (see Table 1.4 and Table 1.5). Furthermore, constitutively GUS-expressing *P. infestans* transformants have been used for quantification of plant resistance levels (Kamoun *et al.*, 1998a), for monitoring the pathogen's progress *in planta* (van West *et al.*, 1998), for promoter studies (McLeod *et al.*, 2004; Tani and Judelson, 2006), and for the localisation of proteins during development (Ah Fong and Judelson, 2003; Cvitanich and Judelson, 2003a; Tani and Judelson, 2006).

Traditional gene disruption methods, invaluable for the functional characterisation of genes, are severely hampered in oomycetes due to their diploid nature. An alternative strategy, gene silencing, which leads to a knock-down of protein expression mediated by introduced RNA sequences that facilitate the destruction of endogenous mRNA

transcripts, has been used successfully in two *Phytophthora* species (Table 1.5). The endogenous *P. infestans* elicitin *inf1* was silenced after introduction of constructs carrying the sense and antisense coding region of the gene (Kamoun *et al.*, 1998b; van West *et al.*, 1999a). Furthermore, transformants proved to be stable over a series of subcultures (van West *et al.*, 1999a). More recently, the genes *Pigpa1*, encoding a subunit of the heterotrimeric G-protein, and *Pibzp1*, a transcription factor gene, were silenced successfully (Latijnhouwers *et al.*, 2004; Blanco and Judelson, 2005). Zoospores of both transformants showed abnormal swimming behaviour and reduced virulence upon encounter with host plants. Silencing of the cellulose-binding elicitor lectin (CBEL) gene in *P. nicotianae* led to cell-wall thickening of hyphae and impaired adhesion to cellulosic substrate, however, no reduction in virulence was observed (Gaulin *et al.*, 2002). In a recent study, Whisson *et al.* (2005) reported transient silencing of the marker gene *gfp* and two endogenous genes (*inf1* and *cdc14*) in *P. infestans*, mediated by the introduction of double-stranded RNA molecules.

Taken together, these results show that functional analysis of genes in *Phytophthora* species by means of transformation is feasible. However, the limited number of published reports over the past 15 years clearly demonstrate that transformation protocols have to be further adapted to improve transformation rates, and thus generate a real high-throughput tool for *Phytophthora* genomics.

Table 1.4 Development of stable transformation protocols and vector systems in *Phytophthora* species (non-gene specific)

				Selection Marker			Marker gene				
Organism	**Method**		**Vector**	**P**	**G**	**T**	**P**	**G**	**T**	**Result**	**Reference**
P. infestans	Lipofectin-PEG	Pro	pHAMT34H	*ham34*	***hpt***	*ham34*		-		✓	Judelson *et al.*, 1991
			pHAMT34N	*ham34*	***nptII***	*ham34*		-		✓	
			pTH209	*hsp70*	***nptII***	*ham34*		-		✓	
			pTH210	*hsp70*	***hpt***	*ham34*		-		✓	
	Lipofectin-PEG	Pro	pTH207				*hsp70*	***gus***	*ham34*		Judelson, 1993[a]
			pTH209	*hsp70*	***nptII***	*ham34*					
			pTH210	*hsp70*	***hpt***	*ham34*					
			pTH211	*hsp70*	***spt***	*ham34*					
			pHAMT34H	*ham34*	***hpt***	*ham34*					
			p35S-HAM				*35S*	***gus***	*ham34*		
			pHAMT35G				*ham34*	***gus***	*ham34*		
			pDHAMT35L				*ham34*	***lacZ***	-		
	Lipofectin-PEG	Pro	pTH207				*hsp70*	***gus***	*ham34*	✓	Judelson *et al.*, 1993b
			pHAMT35G				*ham34*	***gus***	*ham34*	✓	
			pDHAMT35G				*ham34*	***gus***	*ham34*	✓	
			pPMA-HAM		b)		*act*	***gus***	*ham34*	✓	Judelson and Whittaker, 1995
			pPAC-HAM				*actA*	***gus***	*ham34*	✓	Judelson, 1997
			pPUB-HAM				*ubi3R*	***gus***	*ham34*	✓	
			pELI-HAM				*para1*	***gus***	*ham34*	✓	
	A. tumefaciens	Zoo	pNPTII	*hsp70*	***nptII***	*ham34*				✓	Vijn and Govers, 2003
			pNPTII-GUS	*hsp70*	***nptII***	*ham34*	*ham34*	***gus***	*ham34*	✓	

Table 1.4 Development of stable transformation protocols and vector systems in *Phytophthora* species (non-gene specific) - continued

Organism	Method		Vector	Selection Marker P	Selection Marker G	Selection Marker T	Marker gene P	Marker gene G	Marker gene T	Result	Reference
P. infestans	Microprojectile bombardment	Myc	pHAMT35G		-		*ham34*	***gus***	*ham34*	✓	Cvitanich and Judelson, 2003b
		Spo	pTH209	*hsp70*	***nptII***	*ham34*				✓	
		Zoo	pTH209-35G	*hsp70*	***nptII***	*ham34*	*ham34*	***gus***	*ham34*	✓	
	Lipofectin-PEG	Pro	p34GFN	*ham34*	***nptII***	*ham34*	*ham34*	***gfp***	*ham34*	✓	Si-Ammour *et al.*, 2003
P. megasperma	Lipofectin-PEG	Pro	pHAMT34H	*ham34*	***hpt***	*ham34*		-		✓	Judelson *et al.*, 1993a
			pTH210	*hsp70*	***hpt***	*ham34*				×	
P. nicotianae	Lipofectin-PEG	Pro	pAHB6		*c)*		*hsp70*	***sgfp***	*nos*	✓	Bottin *et al.*, 1999
P. palmivora	Lipofectin-PEG	Pro	pHAMT35G		*d)*		*ham34*	***gus***	*ham34*	✓	van West *et al.*, 1999b
			pVW2				*ham34*	***sgfp***	*ham34*	✓	
P. brassica	Lipofectin-PEG	Pro	p34GFN	*ham34*	***nptII***	*ham34*	*ham34*	***gfp***	*ham34*	✓	Si-Ammour *et al.*, 2003

Zoo – Zoospores, Pro – Protoplasts, Myc – Mycelia, Spo – Sporangia, P – Promoter, G – Gene, T- Terminator, nptII – neomycin phosphotransferase, hpt – hygromycin phosphotransferase, hsp70 and ham34 – promoter and terminator fragments from the *hsp70* and *ham34* genes of *B. lactucae*, 35S – promoter from cauliflower mosaic virus 35S gene, spt – streptomycin phosphotranspherase, act – promoter fragment from actin gene of *P. megasperma*, actA – promoter fragment from actin gene of *P. infestans*, ubi3R - promoter fragment from ubiquitin gene of P. infestans, para1 – promoter fragment of para1 elicitin gene of P. nicotianae, nos – terminator fragment of a nopaline synthase gene of maize, gus – β-glucuronidase, sgfp – enhanced GFP, lacZ - -galactosidase. a) a wide range of plasmid combinations used in experiments – refer to paper for details, b) co-transformation with pTH209 or pHAMT35H, c) co-transformation with pTH210 and d) co-transformation with pTH209

Table 1.5 Functional analysis of genes by gene silencing or GUS-reporter

Name		Organism	Conclusion	Reference
Gene silencing				
Elicitin	*inf1*	*P. infestans*	INF avirulence factor in interaction between *N. benthamiana* and *P. infestans*	Kamoun *et al.*, 1998b
			trans-acting factor that mediates silencing signal between nuclei	van West *et al.*, 1999a
Cellulose-Binding Elicitor Lectin	*CBEL*	*P. nicotianae*	cell-wall thickening; impaired adhesion and differentiation in contact with cellulosic substrates	Gaulin *et al.*, 2002
Cell division control protein (Phosphatase)	*piCdc14*	*P. infestans*	affects sporangia development	Ah Fong and Judelson, 2003
G-protein β-subunit	*Pigpb1*	*P. infestans*	is required for sporangium formation; effects cdc14 expression	Latijnhouwers and Govers, 2003
G-protein α-subunit	*Pigpa1*	*P. infestans*	affects zoospore swimming behaviour and taxis, and virulence	Latijnhouwers *et al.*, 2004 Dong *et al.*, 2004
Basic leucine zipper protein	*Pibzp1*	*P. infestans*	affects zoospore swimming behaviour , germination and virulence	Blanco and Judelson, 2005
GFP Elicitin Cell division control protein	*gfp* *inf1* *cdc14*	*P. infestans*	transient silencing of *gfp, inf1, cdc14* by introduction of dsRNA molecules	Whisson *et al.*, 2005

Table 1.5 Functional analysis of genes by gene silencing or GUS-reporter - continued

Name		Organism	Conclusion	Reference
GUS reporter (promoter analysis of genes listed)				
In-planta induced protein	*ipiO*	*P. infestans*	expressed in zoospores, cysts, and germinating cysts; predominantly in subapical and vacuolated tips of invading hyphae *in planta*	van West *et al.*, 1998
Cell division control protein (Phosphatase)	*piCdc14*	*P. infestans*	active during sporangia formation, maturation and germination	Ah Fong and Judelson, 2003
Puf-family protein (RNA binding proteins)	*M90*	*P. infestans*	strong expression in oospores (sexual sporulation), sporangia and zoospores (asexual sporulation)	Cvitanich and Judelson, 2003a
β-1,3-glucanase	*Piexo1* *Piendo1*	*P. infestans*	used for promoter studies	McLeod *et al.*, 2004
Nuclear LIM interactor-interaction factor	*PinifS (Pisp14)* *PinifC1-3*	*P. infestans*	activated in late sporulation stage involved in zoosporogenesis	Tani *et al.*, 2005
	PinifC		promoter studies, detection of a 'cold-box' motif	Tani and Judelson, 2006
Protein kinase	*PK3*	*P. infestans*	activated in hyphal tips before sporangia development	Judelson and Blanco, 2005

1.3.2 Characterisation of *Phytophthora* genes

As transformation and its applicability as an analysis tool in *Phytophthora* research is not straight forward due to the afore mentioned factors, in many situations other methods to characterise genes have been employed. Over recent years, *in silico* analysis of DNA and protein sequences has become a vital part of biological research. This includes, for example, sequence alignments, gene pattern and profile analysis, protein structure predictions, phylogenetic analysis and molecular dynamics studies. Bioinformatic applications, which help researchers to perform these *in silico* experiments, are readily available for use in the form of programs or internet web applets. Other laboratory based techniques which can be used to characterise and further elucidate a gene and its potential function are, for example, cDNA and/or BAC library screening, gene expression studies, enzyme assays, component analysis such as amino acids, production of antibodies and immunolocalisation studies, enzyme-linked immunosorbent assays (ELISAs), yeast complementation assays, adhesion assays and proteomics. Some publications in *Phytophthora* research utilising a number of these techniques are listed in Table 1.6.

Table 1.6 A selection of publications that describe the characterisation of genes or gene families without the use of transformation experiments

Organism	Gene(s)	Methods	Reference
P. cinnamomi	Polygalacturanases	Enzyme assays Sequence analysis Phylogenetic analysis Protein structure modelling	Götesson *et al.*, 2002
P. infestans	Kazal-like protease inhibitor	Expression analysis Enzyme assays Prediction of inhibition constants	Tian *et al.*, 2005 Tian and Kamoun, 2005
Phytophthora species	Elicitins	BAC library screening Phylogenetic analysis Protein structure prediction	Jiang *et al.*, 2005 Jiang *et al.*, 2006b
P. nicotianae	PnPMA1 (plasma membrane H^+-ATPase)	BAC library screening Production of antibodies Immunolocalisation Yeast complementation assay	Shan *et al.*, 2006
P. infestans	Glutathione transferase	Enzyme assays Production of antibodies Production of recombinant protein	Bryant *et al.*, 2006
Phytophthora species	Glycosyl hydrolase	Expression analysis Phylogenetic analysis	Costanzo *et al.*, 2006

1.4 Objectives of this study and experimental approach

The objective of the research described in this thesis was the identification and characterisation of two genes obtained during the screening of randomly selected cDNA clones from *P. nicotianae* cDNA libraries, and the localisation of the respective proteins in *P. nicotianae* zoospores.

P. nicotianae was chosen over other *Phytophthora* species, as it represents a good model organism. It displays a wide host range, is an economically important pathogen and produces large numbers of zoospores in culture.

Chapter 3 describes the isolation and characterisation of the V-PPase gene initially isolated from a *P. nicotianae* zoospore cDNA library. The complete cDNA sequence was obtained by 5' RACE (rapid amplification of cDNA ends) and the expression profile of the gene was determined by RNA blot analysis. Pyrophosphatase activity assays were carried out using microsomal fractions of *P. nicotianae* zoospores, and effectiveness of a type-specific inhibitor was tested *in vitro*, followed by *in vivo* experiments. Monoclonal and polyclonal antibodies directed against highly conserved regions of the protein were generated and used to localise the pyrophosphatase protein in *P. nicotianae* zoospores.

Experiments described in Chapter 4 were designed to characterise a small aquaporin gene family in *P. nicotianae*, following the discovery of an aquaporin encoding cDNA clone isolated from a *P. nicotianae* germinated cysts cDNA library. Five aquaporin genes were obtained by BAC subcloning and sequencing. The gene expression profile of all five genes was obtained by northern blot analysis and a modified single-strand conformation polymorphism (SSCP) approach. Polyclonal antibodies were raised against synthetic peptides and localisation studies were carried out using *P. nicotianae* zoospores.

In Chapter 5, genomic and EST database, and phylogenetic analyses of both genes are described.

Chapter 2

Materials and Methods

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Enzymes, chemicals and molecular biology kits

Restriction endonucleases and DNA modifying enzymes were purchased from New England Biolabs ([NEB], Ipswich, MA, USA), Roche Molecular Biochemicals (Basel, Switzerland), and Promega (Madison, WI, USA). Molecular biology grade chemicals, reagents and antibiotics were purchased from Sigma-Aldrich (St. Louis, MO, USA), Invitrogen (Carlsbad, CA, USA), and GE Healthcare (Uppsala, Sweden). Plasmid DNA isolation miniprep and midiprep kits were purchased from QIAGEN (Hilden, Germany). Gel extraction and polymerase chain reaction (PCR) purification kits were obtained from QIAGEN and Promega. Radioisotope [α-^{32}P] dCTP (3000 Ci/mmol) was purchased from GE Healthcare.

2.1.2 Markers

1 kb plus DNA Ladder and 1 kb DNA Ladder were purchased from Invitrogen and NEB, respectively. For RNA gels, either the Promega 0.28–6.58 kb RNA Ladder or the 0.24-9.5 kb RNA Ladder from Invitrogen was used.

For protein gels, Kaleidoscope Precision Plus and Kaleidoscope Prestained markers were obtained from Bio-Rad Laboratories (Hercules, CA, USA).

2.1.3 Disposable materials

90 mm Petri dishes were supplied by Bacto Laboratories (Liverpool, NSW, Australia) and Greiner Bio-One (Solingen, Germany). 150 mm Petri dishes, 96-well microtiter plates, and 0.2 mL microfuge tubes were obtained from Sarstedt (Nümbrecht, Germany). Pipettes tips were purchased from Edwards Instruments Co. (Narellan, NSW, Australia). 1.5 mL and 2 mL tubes were obtained from Eppendorf (Hamburg, Germany) and Sarstedt. 15 mL and 50 mL Falcon tubes were purchased from Becton Dickinson Labware (Franklin Lakes, NJ, USA). Miracloth was purchased from Calbiochem-Behring Corp. (La Jolla, CA, USA).

2.1.4 Antibiotics

The antibiotics ampicillin, tetracycline, kanamycin and chloramphenicol were added to bacterial growth media at final concentrations of 100 µg/mL, 12.5 µg/mL, 50 µg/mL and 12.5 µg/mL, respectively, for propagation of *Escherichia coli* cultures harbouring plasmid DNA with their respective resistance genes. All of these antibiotics were filter sterilised (0.22 µm) and ampicillin, tetracycline and chloramphenicol were stored as 50 mg/mL, 35 mg/mL and 34 mg/mL stock in 70% ethanol at –20°C until use. Kanamycin was stored as a 10 mg/mL stock at 4°C until use.

2.1.5 Bacterial strains

E. coli SOLR™ strain (kanamycin resistant) and XL1-Blue MRF′ strain (tetracycline resistant) (Stratagene Inc., La Jolla, CA, USA) were used as host bacteria for cDNA library mass excision and isolation of individual cDNA clones. Following cDNA library excision the plasmids containing the cDNA clones were transformed into a more stable host, *E. coli* strain DH10β™ (Invitrogen) and cultured using ampicillin resistance selection. The genotypes of all the *E. coli* strains used in this study are presented in Appendix B.1

2.1.6 Antibodies

Antibodies used in this study are presented in Appendix C.

2.2 Methods - Molecular Biology: Preparative DNA techniques

Detailed recipes for the buffers and solutions used in this study are presented in Appendix A. A variety of bacterial growth media were used during this study and detailed recipes are given in Appendix B.2. The *Phytophthora* culture media recipes are presented in Appendix D.

2.2.1 Preparation of plasmid DNA

Plasmids were isolated from a 5 mL overnight culture of *E. coli* grown in Luria-Bertani (LB) broth + 100 µg/mL ampicillin using the QIAprep® Spin Miniprep kit (QIAGEN). 5 mL overnight bacterial cultures were pelleted by centrifuging at 16,000 x *g* for 1 minute (min) in an Eppendorf micro centrifuge. The bacterial pellet was resuspended in 250 µL of Buffer P1 using a 1 mL pipette tip, 250 µL of Buffer P2 was added to the mixture and the tube was gently inverted 4-6 times to mix the contents. Then 350 µL of Buffer N3 was

added and the suspension was mixed immediately but gently by inverting the tubes four to six times. The mixture was centrifuged for 10 min at 16,000 x *g* and the supernatant was transferred into a QIAprep spin column. The spin column was inserted in a 2 mL collection tube and centrifuged for 1 min at 16,000 x *g*. The flow-through in the collection tube was discarded, the QIAprep spin column was washed with 0.5 mL of wash Buffer PB and centrifuged for 1 min at 16,000 x *g*. The flow-through in the collection tube was discarded and the QIAprep spin column was washed with 0.75 mL of wash Buffer PE and centrifuged for 1 min at 16,000 x *g*. The flow-through in the collection tube was discarded and the QIAprep spin column was centrifuged again for 1 min to remove residual Buffer PE from the column. The QIAprep spin column was transferred to a sterile 1.5 mL microfuge tube. To elute the plasmid DNA, 50 µL of distilled water (dH_2O) was added to the filter of the QIAprep column, allowed to stand for 1 min and then centrifuged for 2 min at 16,000 x *g*. The DNA was used for restriction digests, PCR applications and sequencing reactions. The DNA was separated on agarose gels (2.3.1), examined under an ultra violet (UV) transilluminator and photographed using the GelDoc Imaging system as described in Section 2.3.2. Plasmid DNA quality and concentration was determined according to protocols in Section 2.3.3.

In order to obtain larger amounts of plasmid DNA (up to 200 µg/Midi-prep or 750 µg/Maxi-prep) the Qiafilter Midi and Maxi Kit (QIAGEN) was used following manufacturer's instructions.

2.2.2 Preparation of genomic DNA

Hyphal tissue of *P. nicotianae* H1111 and *P. sojae* 1167 grown on V8 agar were harvested and snap frozen in liquid-N_2 according to the protocol described in Section 2.6.1. For the isolation of genomic DNA, the frozen tissue was ground to a fine powder in liquid-N_2 and transferred to 50 mL Falcon tubes containing 20 mL of *Phytophthora* genomic DNA extraction buffer. One Falcon tube was used for hyphae harvested from at least ten plates which represented ~ 6 g of tissue. The suspension was mixed thoroughly by vortexing and then divided into 10 mL volumes in 2 x 30 mL Oak ridge tubes. An equal volume of phenol:chloroform (1:1) was added to the Oak ridge tubes and mixed thoroughly. The resulting mixture was centrifuged at 5,600 x *g* for 10 min at 4°C. After centrifugation the upper phase of the mixture was pipetted to fresh 30 mL Oak ridge tubes and re-extracted with an equal volume of phenol:chloroform. The mixture was centrifuged as described

above. The phenol:chloroform extraction and centrifugation step was repeated once more and then the upper aqueous layer was transferred to a 30 mL Corex tube. A 0.1 volume of 3 M NaOAc, pH 4.8, and a 0.6 volume of isopropanol was added, the suspension was incubated on ice for 30 min and then centrifuged for 10 min at 10,000 x *g* at 4°C. The resulting pellet was washed with 70% ethanol, dissolved in 500 µL of Tris/EDTA (TE) buffer supplemented with 100 µg/mL RNAse A (Promega) and incubated at 37°C for 30 min. The mixture was extracted once with phenol:chloroform and centrifuged at 16,000 x *g* in an Eppendorf micro centrifuge for 10 min at 4°C. The upper aqueous layer was transferred to a fresh 2 mL microfuge tube, washed with 0.1 volume 3 M NaOAc, pH 4.8, and 2 volumes of absolute ethanol and incubated on ice for 30 min. The genomic DNA was pelleted by centrifugation at 16,000 x *g* in an Eppendorf micro centrifuge for 15 min at 4°C. The DNA pellet was washed once with 70% ethanol, briefly air-dried and dissolved in ddH_2O. The DNA quality and concentration was determined according to protocols described in Sections 2.3.1 to 2.3.3.

2.2.3 Concentration of DNA

Two volumes of cold 100% ethanol or 0.7 volumes of isopropanol plus 1/10 volume of 3 M sodium acetate, pH 5.2 were added to the nucleic acid solution to be precipitated, followed by incubation for 10 min on ice. To recover the precipitated DNA, samples were centrifuged at 16,000 x *g* for 30 min at 4°C. The pellet was rinsed with 70% ethanol and subsequently air dried at room temperature (RT) after re-centrifugation.

2.2.4 Phenol/chloroform extraction

Proteins were removed from an aqueous solution of DNA by extracting with an equal volume of phenol (saturated in 1 M Tris-HCl, pH 7.5). The sample was completely mixed by inverting the tube several times and the phases were separated by centrifugation at 16,000 x *g* for 5 min at RT. The upper, aqueous layer was transferred to a new tube and mixed with an equal volume of phenol and chloroform. Repeated mixing and spinning yielded a new aqueous phase, which was eventually mixed with an equal volume of chloroform to remove traces of phenol. After the phases were separated, the aqueous phase was pipetted into a fresh tube and DNA was concentrated by isopropanol precipitation as described in Section 2.2.3.

2.2.5 DNA extraction from agarose gels

Following agarose gel electrophoresis (2.3.1) and ethidium bromide (EtBr) staining, the DNA band requiring purification was excised from the agarose gel using a clean scalpel and placed in a clean microfuge tube. DNA bands were isolated from agarose gel using either the QIAquick Gel Extraction Kit (QIAGEN) or the UltraClean™15 DNA Purification Kit (MoBio Laboratories, Carlsbad, CA, USA) following manufacturers' instructions.

2.2.5.1 QIAquick Gel Extraction Kit

QXI Buffer 1 was added to the tube containing the DNA fragment at 3:1 ratio of buffer to the weight of the gel and incubated at 50°C for 10 min to dissolve the agarose. During the incubation period the tube was vortexed several times to solubilise the agarose. The mixture was loaded onto a QIAquick spin column and centrifuged for 2 min at 16,000 x *g* in an Eppendorf micro centrifuge. The DNA bound to the QIAquick spin column was washed with 800 µL of ethanol containing Buffer PE and centrifuged for 2 min at 16,000 x *g*. The QIAquick spin column was removed and placed in a fresh microfuge tube and centrifuged again to remove residual wash buffer. The DNA from the QIAquick spin column was eluted in 50 µL double-distilled water (ddH_2O) after centrifuging for 2 min at 16,000 x *g*. The DNA quality and concentration was determined according to protocols described in Sections 2.3.1 to 2.3.3.

2.2.5.2 UltraClean™15 DNA Purification Kit

Three volumes of ULTRA SALT was added to the gel and incubated at 55°C until the gel was completely melted. ULTRA BIND was vortexed for 30 seconds (sec) to homogenise the DNA binding silica matrix and then added to the solubilised agarose in a ratio of 6 µL per expected µg DNA. The mixture was incubated at RT for 5 min. During the incubation period the tube was carefully mixed several times to allow for good contact between DNA and the silica beads. Following incubation the tube was centrifuged for 5 sec in an Eppendorf micro centrifuge at 16,000 x *g*. The supernatant was discarded and the pellet resuspended in 1 mL of ULTRA WASH by vortexing for 5-10 sec or soaking for 5 min (for DNA fragments larger than 15 kb). The suspension was then centrifuged for 5 sec and the supernatant was discarded. To completely remove all traces of the washing buffer the tube was re-centrifuged for 5 sec, the remaining supernatant was aspirated with a narrow tip pipet and the pellet was allowed to air-dry for several

minutes. For elution of the DNA, a volume twice of that of the ULTRA BIND matrix was added to the pellet and then resuspended by pipetting until homogenous. After an incubation period of up to 5 min the suspension was centrifuged for 1 min at 16,000 x *g*. The supernatant was recovered and transferred to a fresh tube. The DNA quality and concentration was determined according to protocols described in Sections 2.3.1 to 2.3.3.

2.2.6 DNA probe synthesis

DNA templates for use in RNA and DNA blots were labelled with either [α-^{32}P]-dCTP using the Megaprime DNA Labelling Kit (GE Healthcare) or with Digoxigenin-11-dUTP using the DIG-High Prime DNA Labelling Kit (Roche Molecular Biochemicals) as described hereafter.

2.2.6.1 Megaprime DNA labelling

The labelling of DNA was done according to manufacturer's instructions using the following volumes:

Component	Volume
Template DNA (PCR product) (200 ng)	1-3 µL template DNA
Primer solution	4 µL
Sterile ddH_2O	16.8-14.8 µL (volume to here should be 21.8 µL)
Unlabelled dNTP's	3.2 µL each = 9.6 µL total (dATP, dGTP, dTTP)
Reaction buffer	4 µL
Radiolabelled [α-^{32}P]-dCTP	3 µL
Klenow Enzyme (1U)	1.6 µL
Total Volume	40 µL

(The template DNA and primer were denatured by boiling for 5 min and then cooled one ice for 1 min before addition of other components)

The final reaction volume was mixed by gentle pipetting and centrifuged for 5 sec to bring the contents to the bottom of the tube and incubated at 37°C for 30 min after which 60 µL of TE buffer, pH 8.0, was added to bring the volume up to 100 µL. Unincorporated nucleotides were removed from this reaction mix by filtering through a Sephadex G50 (GE Healthcare) column prepared using a glass Pasteur pipette. The ^{32}P-labeled DNA was denatured by boiling for 5 min, chilled on ice for 1 min and then added carefully to the Pre-hybridisation buffer as described in the Section 2.3.5.

2.2.6.2 DIG-High Prime DNA Labeling

1 µg template DNA was mixed with ddH_2O to a final volume of 16 µL in a reaction tube and DNA was denatured by heating in a boiling water bath for 10 min and quickly chilled on ice. DIG-High Prime was thoroughly mixed and 4 µL were added to the denatured DNA. The suspension was mixed and briefly centrifuged and then incubated at 37°C for 20 hours (h). The reaction was stopped by adding 2 µL of 0.2 M ethylenediamine tetra-acetic acid (EDTA) (pH 8.0) and heating to 65° C for 10 min. The DIG-labelled probe was denatured (~ 25 ng/mL) by boiling for 5 min, rapidly cooled on ice and then added to the pre-hybridisation buffer (2.3.5.2). The remaining probe was stored at -20°C.

2.2.7 cDNA library screening

A cDNA expression library using mRNA isolated from *P. nicotianae* zoospores was constructed by Dr D. Škalamera in λZAPII according to the supplier's manual (Stratagene, La Jolla, CA, USA).

2.2.7.1 Cell preparation

E. coli (strain XL1-Blue MRF′) was grown in a sterile 50 mL Falcon tube containing 20 mL of LB broth, 0.2% maltose and 10 mM $MgSO_4$ overnight at 37°C with shaking at 250 revolutions per min (rpm). The cells were centrifuged at 1000 x *g* for 5 min at RT, the supernatant was decanted and the pellet was resuspended in 10 mL of sterile 10 mM $MgSO_4$. These cells were viable for at least 1 week at 4°C.

2.2.7.2 First round of screening

Approximately 4 x 10^4 plaque forming units (pfu) were mixed with 600 µL of XL1-Blue MRF' cells in 10 mM $MgSO_4$ (Optical density $[OD]_{600}$=0.5) in large sterile culture tubes and incubated at 37°C for 30 min to allow attachment of phage to the bacterial cells. 6.5 mL of NZY top agar, molten and kept warm in a 50°C water bath was added to each tube containing the λ phage and XL1-Blue MRF' bacterial cells, and mixed by rolling the tubes. This mixture was poured onto pre-warmed 150 mm NZY agar plates, the top agar was allowed to set and the plates were incubated at 37°C overnight in an inverted position to prevent condensation from dripping on the agar surface. The plates containing a nearly confluent growth of plaques were removed from the incubator and placed at 4°C and allowed to cool for at least 30 min. Two sets of six 137 mm Hybond N^+ membranes (GE Healthcare) were cut and numbered accordingly. After cutting

orientation identifying marks on the membranes, a membrane was overlaid on each plate by gently bending it in the middle and placing onto the surface of the agar. The membranes were allowed to come into contact with the plaques on the agar surface and were left in place for at least 1 min. The orientation of the membrane on each plate was matched with marks made on the same plate. The membranes were removed from the plates with a pair of forceps and placed sequentially, DNA side up, on pieces of Whatman 3MM (Middlesex, UK) filter paper saturated with Denaturing buffer for 2 min followed by Neutralisation buffer for 2 min and 5x SSC for 5 min. After this treatment, the membranes were air-dried and DNA was fixed to the membrane by baking at 80°C for 2 h sandwiched between two sheets of dry 3MM Whatman filter paper. The second set of membranes was prepared as described for the first set. The baked membranes were washed in 2x SSC, transferred to a bowl containing 400-500 mL Colony wash solution and incubated for 30 min with gentle agitation to remove any cellular debris. The membranes were placed in separate hybridisation bottles containing 50 mL pre-hybridisation solution and incubated for 2 h at 65°C in a Hybaid hybridisation oven (Thermo Electron Corp., Waltham, MA, USA). The next day membranes were washed twice in 2x SSC/0.1% sodium dodecyl sulphate (SDS) at 65°C for 15 min each, followed by two washes for 15 min each in 0.2x SSC/0.1% SDS at 65°C. Images were acquired as described in Section 2.3.5.3.1.

2.2.7.3 Plaque purification

Putative positive pfu were picked from the first round of screening and used to infect fresh XL1-Blue MRF' cells. An agar plug containing the positive phage (and possibly contaminating phage) was removed with the large end of a pasteur pipette and the phage eluted overnight into 500 µL SM buffer containing 20 µL chloroform at 6-8°C. The phage stock was re-screened on Petri dishes with a diameter of 90 mm until pure plaques were obtained as described in the previous section.

2.2.7.4 Phagemid excision

After purification, the phagemid was excised and sequenced. In this procedure, XL1-blue MRF' and SOLR cells were grown in LB broth supplemented with 0.2% maltose and 10 mM $MgSO_4$ at 30°C overnight. The cells (OD_{600}=1.0) were collected by centrifugation at 1000 x *g* and resuspended in 10 mM $MgSO_4$. In a sterile 15 mL Falcon tube, 200 µL XL1-Blue MRF' cells, 250 µL phage stock and 1 µL ExAssist® helper phage (Stratagene)

were incubated at 37°C for 15 min. Three mL of LB broth (pre-warmed at 37°C) was added to the above reaction mixture and incubated at 37°C at 250 rpm for 3 h. The suspension was heated to 70°C for 20 min and the cell debris removed from the excised pBS phagemid, packaged as filamentous phage particles, by centrifugation for 15 min at 1000 x *g* at RT. The supernatant was decanted into a sterile 15 mL Falcon tube and stored at 4°C. The excised phagemids were viable for at least 2 months at this storage temperature. The excised phagemid was grown on SOLR cells (10 or 100 µL phagemid solution plus 200 µL SOLR cells per dilution) on LB agar containing 50 µg/mL ampicillin after incubation of host cells and phagemid for 15 min at 37°C. The colonies that appeared on the agar contained the pBS double stranded phagemid. Single colonies of *E. coli* SOLR cells containing the phagemid were grown in 5 mL LB broth containing 75 µg/mL ampicillin for at least 5 h on a shaker. DNA was extracted according to Section 2.2.1.

2.2.8 Isolation of genomic clones from a BAC library

Genomic clones were isolated from a large-insert BAC library which was constructed using nuclear DNA from *P. nicotianae* strain H1111 (Shan and Hardham, 2004). This entire BAC library was pooled on a single filter and each lane on the BAC pooled filter represented 1536 bacterial colonies from 4 x 384-well plates which were spotted on a single 10x10 cm filter. Seven filters representing 28 x 384-well plates comprised the whole genomic library.

2.2.8.1 Screening

Pre-hybridisation and hybridisation of BAC colony filters was performed according to protocols described in Section 2.3.5.2. Plasmid DNA from cDNA clones used as probes in hybridisation reactions was isolated using the QIAprep Spin Kit miniprep protocol as described in Section 2.2.1. Inserts were released from cDNA clones by *Eco*RI and *Xho*I double digestion (2.3.6), purified by gel electrophoresis (2.2.5) and labelled with ^{32}P as described in Section 2.2.6.1.

2.2.8.2 Isolation of genomic clones

Putative positive signals on BAC filters were picked for DNA isolation. Genomic clones were grown in 2 mL LB broth plus chloroamphenicol (12.5 µg/mL) overnight with shaking at 280 rpm at 37°C. The following day the bacterial pellet was recovered after centrifugation at 16,000 x *g* for 30 sec. The cells were suspended in 100 µL Solution

I/Resuspension buffer, 200 µL Solution II/Lysis buffer was added to the suspension and mixed gently by inverting the tube several times. The tubes were placed on ice for 5 min, 100 µL of Solution III/Neutralisation buffer was added and the suspension was mixed gently by inverting the tubes several times. The tubes were incubated on ice for 10 min and the suspension was centrifuged for 10 min at 16,000 x *g* at RT. The supernatant was recovered after the centrifugation and was extracted once with phenol:chloroform according to Section 2.2.4. The DNA quality and concentration was determined according to protocols described in Sections 2.3.1 to 2.3.3.

2.3 Methods – Molecular Biology: Analytical DNA techniques

2.3.1 Agarose gel electrophoresis

Plasmid DNA were size-fractionated in horizontal submerged TAE (Tris-acetate/EDTA)-agarose gels according to protocols described in Sambrook and Russell (2001). Agarose 0.8 to 2.0% was dissolved in 1x TAE buffer. One fifth volume of DNA loading buffer was added to the DNA before loading onto the gel. The DNA was separated at 2-15 V/cm for 45-180 min and a 1 kb DNA Ladder (NEB or Invitrogen) at 500 µg/mL was routinely run on the agarose gels to determine the size and give an estimate of the concentration of the unknown DNA on these gels. The 3 kb band in the 1 kb NEB DNA Ladder contained 125 ng DNA when run at 500 µg/mL, whereas the 1636 bp band of the 1 kb Invitrogen DNA Ladder represented 10% of the mass applied to the gel.

2.3.2 Visualisation and imaging of nucleic acids

Following agarose electrophoresis nucleic acids were stained in 0.5 µg/mL EtBr in ddH_2O for 15-20 min and destained for 15-30 min in 1x TAE buffer. The DNA was visualized by illumination on a UV transilluminator (340 nm) and photographed using the GelDoc Imaging system (Bio-Rad Laboratories).

2.3.3 Quantification of nucleic acids

As an alternative to agarose gel quantification of DNA (2.3.1), a UV spectrophotometer (Ultrospec III, Pharmacia - GE Healthcare) was used also to quantify plasmid and genomic DNA, and RNA preparations. Absorbance at 260 nm was used to measure the concentration of DNA and RNA as described by Sambrook and Russell (2001). It was assumed that an OD of 1.0 corresponds to 50 µg/mL of double-stranded DNA (dsDNA).

The presence of proteins and other contaminants was estimated by measuring the $A_{260nm/280nm}$ ratio. A ratio of between 1.8 to 2.0 indicated that the DNA preparation was free from contaminating proteins. RNA was quantified using the same procedure. It was assumed that an OD of 1.0 corresponded to 40 µg/mL of RNA. The presence of proteins and other contaminants was estimated by measuring the $A_{260/280\ nm}$ where a ratio of between 1.8-2.0 was taken to indicate clean preparations.

2.3.4 PCR

PCR was used to analyse expression levels of genes after reverse transcription of mRNA to cDNA, to screen bacterial colonies for the presence of linker-insertion after cloning, and to amplify fragments for cloning and screening purposes. The amplification of DNA fragments by PCR was performed in the MJ Research PTC-200 cycler (Waltham, MA, USA) as described below.

The Taq DNA Polymerase and 10x PCR Buffer (Invitrogen) were used in following experiments. A 10 mM mixture of the deoxyribonucleotide triphosphates (dNTPs) dATP, dTTP, dGTP and dCTP (Invitrogen) was made up and stored at -20°C in aliquots. Alternatively, a 2x Master Mix (Promega) which contains a mixture of Taq DNA Polymerase and deoxynucleotides was used. Mg^{2+} and template DNA concentration as well as the annealing temperature varied in individual sets of reactions. Generally, a PCR reaction was carried out in 0.2 mL thin-walled tubes, containing 1/10 volume of 10x PCR Buffer, 1/10 volume of 10x nucleotide solution, 0.6 µM primer, 1-4 units (U) enzyme and 20-100 ng template DNA. Alternatively, bacterial cells were used directly as a template (colony PCR). The final volume was made up to 20 µL with sterile ddH_20.

Reactions were performed under the following general conditions:

Denaturation step	95°C	2 min
Cycling program	[25-35 cycles]	
Denaturation	95°C	30 sec
Annealing	56-62°C	30 sec
Extension	72°C	30-45 sec
Final extension	72°C	7 min

Amplification of PCR products was monitored by gel electrophoresis (2.3.1).

In some instances a gradient PCR was carried out to determine the optimal annealing temperature and magnesium concentration for a given set of primers. Gradient PCR allows testing of a range of temperatures in one experiment, as a gradient is built up across the thermoblock.

2.3.5 Southern blot analysis

Approximately 20 µg of genomic DNA was digested with the appropriate restriction enzyme (2.3.6) in a final volume of 30 µL for at least 4 h at 37°C and separated on a 1.0% agarose gel using a large gel casting apparatus in 1x TAE buffer over night at 20 volts (V). After electrophoresis the gel was stained for 30 min in EtBr and photographed with a fluorescent ruler using the GelDoc Imaging system (2.3.2).

In preparation for transfer of the genomic DNA from the gel onto the Hybond N^+ membrane the genomic DNA in the gel was depurinated in 500 mL Depurination buffer for 10 min with gentle shaking after which the acid solution was decanted and the gel was washed in ddH_2O for 5 min. The gel was then incubated in 500 mL of Denaturing buffer for 20 min with gentle shaking, washed in ddH_2O for 5 min and incubated in 500 mL of Neutralisation buffer for 30 min with gentle shaking. A Hybond N^+ nylon membrane cut to the size of the gel was soaked in sterile ddH_2O for 1 min followed by 20x SSC for 5 min. Several sheets of Whatman 3MM paper (1 cm wider than the gel) and one piece ~30 cm in length for use as a wick in the transfer apparatus were cut to size.

2.3.5.1 The transfer process

For transfer of DNA from gel to the membrane the steps described in 'Current Protocols for Molecular Biology' - Section 2.9 (Ausubel *et al.*, 1993) were used to set up the transfer apparatus and the set up was left overnight to complete the transfer process (Figure 2.1). The following day the transfer apparatus was dismantled and the gel was recovered and stained with EtBr to confirm the transfer of DNA onto the membrane. The membrane was sandwiched between two sheets of Whatman 3MM paper and baked for 2h at 80°C.

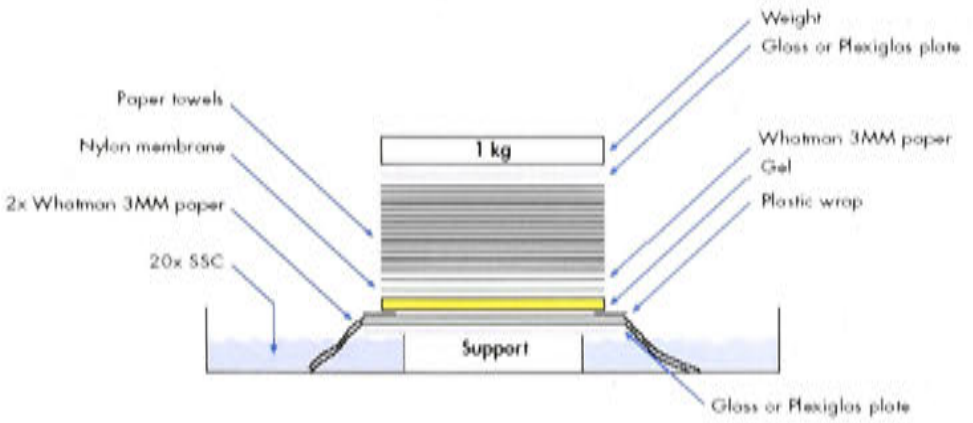


Figure 2.1 The upward capillary transfer set up used for Southern blots in this study. (Source: Ausubel *et al.*, 1993)

2.3.5.2 *Pre-hybridisation and hybridisation*

The membranes were pre-hybridised in 10 mL pre-warmed Hybridisation buffer at 65°C for at least 2 h to suppress non-specific absorption of the probe. After addition of ^{32}P/DIG-labelled DNA probes to the Hybridisation buffer, the membranes were hybridised overnight at 65°C. The next day membranes were washed twice in 2x SSC/0.1% SDS at 65°C for 15 min each, followed by two washes for 15 min each in 0.2x SSC/0.1% SDS at 65°C.

2.3.5.3 *Image acquisition*

2.3.5.3.1 *^{32}P-labelled probes*

Membranes were exposed to phosphor-imaging screens (Molecular Dynamics, Sunnyvale, CA, USA) for 2-72 h depending on expected signal intensity. The hybridisation signals were captured using a PhosphorImager (Molecular Dynamics) and analysed using ImageQuant 5.0 (Molecular Dynamics). The membranes for DNA blots were exposed to phosphor-imaging screens for 2-24 h before scanning, while the membranes for RNA blots were exposed for 48-72 h before scanning.

In order to remove one probe and re-hybridise with another probe (referred to as 'stripping'), membranes were incubated in boiling Stripping buffer and the solution was allowed to cool to RT. To confirm the removal of the first probe the membrane was exposed to a phosphor-imaging screen overnight.

2.3.5.3.2 *DIG-labelled probes*

Membranes were rinsed in Washing buffer for 1-5 min and then incubated in 100 mL Blocking buffer for 30 min at RT. The anti-DIG-alkaline phosphatase (AP) antibody was centrifuged for 5 min at 6,700 x *g* prior to use and then diluted 1:5000 in Blocking buffer. The Blocking buffer was discarded and replaced by 20 mL antibody solution. Following an incubation period of 30 min the membranes were washed twice for 15 min in 100 mL Washing buffer and then equilibrated in 20 mL Detection buffer for 2-5 min. Band development was achieved by addition of the AP substrate (FastRed® tablet, Sigma-Aldrich). The colour reaction was stopped by removing the developing solution and rinsing the membranes in two changes of ddH_2O. The membranes were dried between filter paper.

2.3.6 Restriction endonuclease digestion of DNA

Plasmids, BAC, λ phage and genomic DNA were digested by incubation with 1-2 U of restriction enzyme per 0.5-1.0 µg of DNA in the recommended restriction enzyme buffers supplied by NEB. Restriction digests were carried out for 1 to 16 h depending on the concentration of DNA to be digested at the recommended temperature. Restriction digests of DNA with multiple enzymes were carried out simultaneously if the restriction enzyme buffers were compatible. When different reaction buffers were required, the DNA was digested first with one restriction endonuclease and the buffer was adjusted accordingly to accommodate the second restriction endonuclease. After restriction digests the DNA was cleaned and extracted from agarose gel using a gel extraction kit (2.2.5) before used in cloning or transformation experiments.

2.3.7 Cloning

2.3.7.1 Converting 3'- and 5'-overhangs to blunt ends

When no suitable restriction sites were found in plasmid and insert, blunt end cloning was carried out. Conversion of 3'- and 5'-overhangs was performed by either using the T4 DNA- or Klenow Polymerase (NEB). Both enzymes were used to fill-in 5'-protruding ends with dNTPs or to degrade 3'-overhangs. For 3'-overhang removal usage of T4 DNA Polymerase was preferred because of stronger 3' exonuclease activity.

For blunting, the DNA was provided with 1/10 volume of the appropriate buffer, 100 µM of each dNTP and 5 U of enzyme per µg of DNA. Reactions were incubated for 1 h at 37°C and subsequently purified by using either the UltraClean DNA Purification Kit or the QIAquick Gel Extraction Kit according to Section 2.2.5.

2.3.7.2 Dephosphorylation of linearised plasmid DNA

Calf Alkaline Phosphatase ([CAP], NEB) treatment was used to remove 5'-phosphate overhangs from the pBS vector after restriction digestion and before ligation in cases where a single restriction enzyme site was used for cloning. The phosphatase treatment prevented preferential self-ligation of the vector backbone and facilitated high-efficiency vector plus insert ligations. One unit of CAP enzyme (1U/µL) and 5 µL of 10x CAP buffer were added to 20 µL (0.3-0.5 µg) of restriction digested vector in a final volume of 50 µL and incubated at 37°C for 30 min, after which the CAP enzyme was inactivated by heating at 65 °C for 10 min and removed from the reaction by phenol:chloroform (1:1

ratio) extraction (2.2.4). The DNA was recovered from the supernatant by ethanol precipitation (2.2.3).

2.3.7.3 Ligation

To ligate fragments, 1 U of T4 DNA Ligase (Promega) was added to a 20 µL final reaction volume containing 1x Ligase buffer and 1:1 or 1:3 ratio of vector:insert. Ligation reactions were carried out depending on the nature of the ends. Sticky end ligations were incubated for 1 to 2 h at RT, blunt end ligations either for 4 h at RT or overnight at 16°C depending on the size of vector and insert. An aliquot was run on an agarose gel (2.3.1) to confirm successful ligation before transformation into *E. coli* (2.6.4). In addition, several control ligations were performed to test the efficiency of the ligation reaction and the quality of the vector. These usually included parallel ligations in the absence of insert DNA to determine the background clones arising from self-ligation of an inefficiently dephosphorylated vector.

2.4 Methods – Molecular biology: RNA techniques

2.4.1 Total RNA isolation using TRIzol® reagent

Standard procedures were followed during the handling of RNA to prevent degradation (Sambrook and Russell, 2001). 0.1% diethyl pyrocarbonate (DEPC) treated water was used to prepare all RNA solutions. Glassware was baked at 220°C for 4 h and only aerosol barrier tips supplied as RNase-DNAse free were used. Total RNA from each developmental stage was extracted using TRIzol® reagents (Invitrogen).

Tissue was ground in liquid-N_2 using a mortar and pestle. 100 mg of tissue was added to 1 mL of TRIzol® reagent in a 2 mL microfuge tube and mixed by vortexing for 1 min. The insoluble material from the homogenate was removed by centrifugation at 12,000 x *g* for 10 min at 4°C. The resulting pellet contained cellular membranes, polysaccharides and high molecular weight DNA. The supernatant contained total RNA. The cleared supernatant was transferred to a fresh tube for phase separation using chloroform. The RNA containing supernatant was incubated for 5 min at 22°C in a water bath then 0.2 mL of chloroform was added to the supernatant and mixed by inverting the tube several times and then incubating at 22°C in a water bath for a further 5 min. The mixture was centrifuged at 12,000 x *g* for 15 min at 4°C. After centrifugation the mixture separated into a lower phenol-phase, an interface and a colourless upper

aqueous phase. Total RNA remained exclusively in the upper aqueous phase. For RNA precipitation the upper aqueous phase was transferred to a fresh tube. The total RNA was precipitated from the aqueous phase by mixing with 0.5 mL of isopropanol, incubation of the sample at 22°C for 10 min and centrifugation at 12,000 x *g* for 10 min at 4°C. The RNA-precipitate formed a gel-like pellet on the side and bottom of the tube. After decanting the supernatant the RNA pellet was washed once with 1 mL of 75% ethanol. The sample was mixed by vortexing and centrifuged at 7,500 g for 5 min at 4°C. After decanting the ethanol and brief air-drying the total RNA was re-dissolved in 50 µL of DEPC-ddH_2O per each 100 mg of tissue used for RNA extraction.

RNA quality and concentration was determined according to protocols in Section 2.3.3. Integrity of the RNA preparation was determined by comparing the relative intensities of the 23S and 18S rRNA bands after the denaturing formaldehyde agarose gel electrophoresis (2.4.4).

2.4.2 cDNA synthesis

Prior to cDNA synthesis by reverse-transcriptase-PCR (RT-PCR), contaminating DNA was eliminated by DNaseI treatment (Promega) as per manufacturer's instructions. Briefly, dissolved RNA was incubated with 1/10 volume of 10x RQ1 Reaction Buffer and 1 U RQ1 DNaseI per 1 µg of RNA for 30 min at 37°C. The reaction was stopped by addition of 1 µL of RQ1 Stop Solution per 10 µL reaction and incubation at 65 °C for 10 min. Treated RNA was used in PCR (2.3.4) or SSCP (2.4.4) experiments. If gel electrophoresis was carried out on RNA samples, a phenol:chloroform extraction (2.2.4) was performed before loading the samples onto a gel.

2 µg of the RNA preparation was used for cDNA synthesis reactions using SuperScript™ III Reverse Transcriptase (RT) (Invitrogen). RNA was primed with 2 pmol gene specific primer (GSP) and a 2.5 mM dNTP mix at 65°C for 5 min and quick chilled on ice. A master mix containing 1/5 volume of 5x First Strand Buffer, 0.1 mM dithiothreitol (DTT) and 40 U RNaseOUT (Invitrogen) was added and the reaction was incubated at 42°C for 2 min. Reverse transcription was carried out by adding 200 U SuperScript™ III RT and incubation at 55°C for 50 min followed by 10 min at 70°C to inactivate the reaction. cDNA obtained was stored at -20°C until use.

2.4.3 5' RACE

5' RACE is a PCR based technique which facilitates cloning of cDNA 5'-ends after partial cDNA sequences have been obtained by other methods. The order of experiments is depicted in the flow-chart in Figure 2.2. cDNA synthesis was carried out according to Section 2.4.2. and the product obtained was purified using the UltraClean DNA Purification Kit as described in Section 2.2.5.2. A-tailing of the cDNA was achieved by adding 1/10 volume of NEB buffer 4, 1 mM ATP and, after incubation for 3 min at 94°C and quick chilling on ice, 20 U of DNA Terminal Transferase (NEB). The reaction was incubated at 37°C for 20 min followed by 70°C for 10 min to inactivate the enzyme. PCR amplification was carried out in two rounds. In the first round, a GSP and an anchor primer which primes the A-tail that was added during the tailing step was used for amplification (Table 2.1). The second round PCR, which was carried out according to Section 2.3.4., used a small aliquot of the first round PCR as template, a nested GSP and shorter anchor primer lacking the T-tail. The final product was assessed by gel electrophoresis (2.3.1), purified with the UltraClean DNA Kit (2.2.5.2), cloned into the vector pGEM-T easy (Promega) and sequenced (2.7.1).

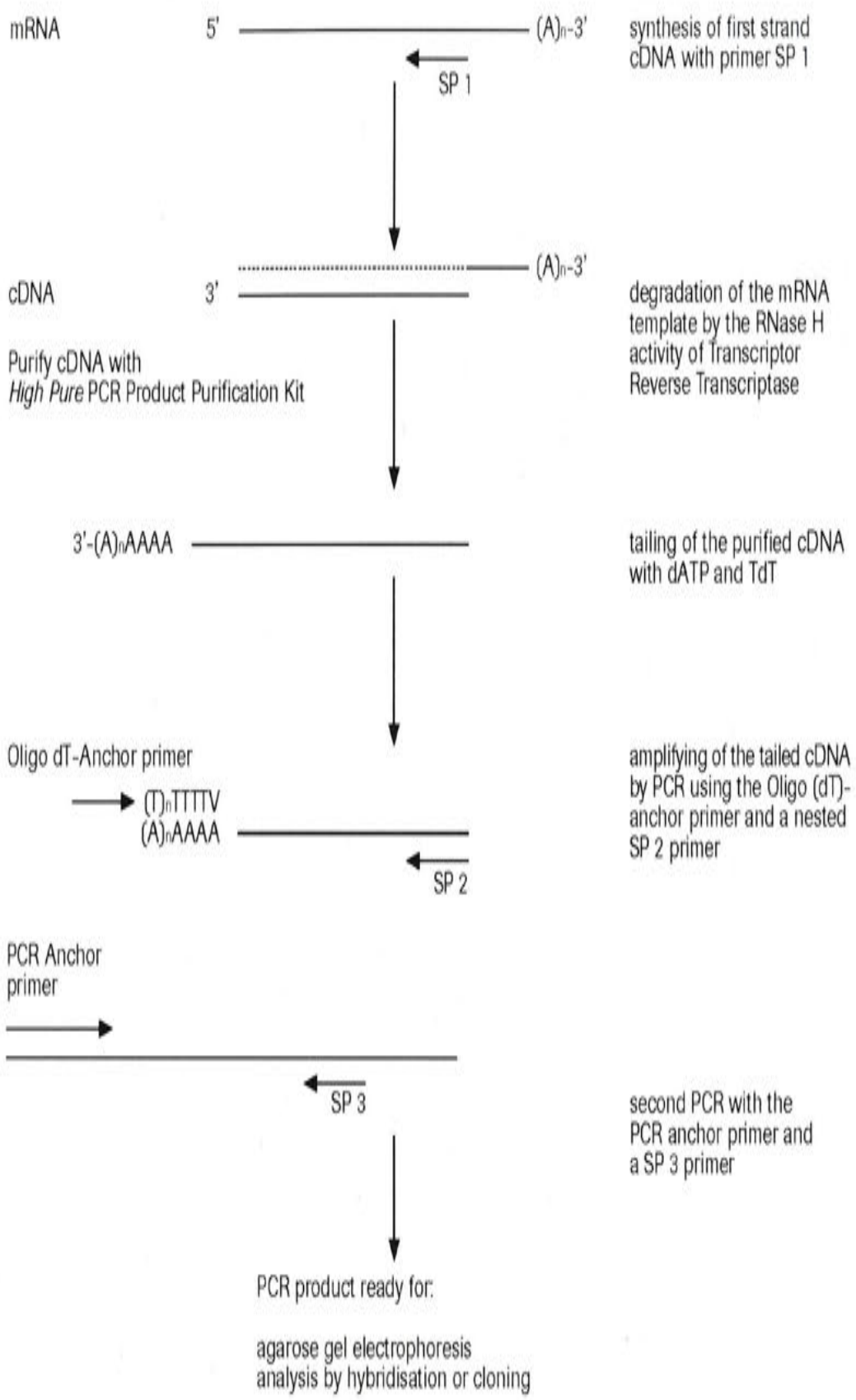


Figure 2.2 5' RACE flow-chart. (Source: Roche, 2005)

Table 2.1 5' RACE – 1st round PCR conditions

Denaturation step	95°C	2 min
Cycling program	[10 cycles]	
Denaturation	95°C	45 sec
Annealing	xx°C	45 sec
Extension	72°C	2 min
	[25 cyles]	
Denaturation	95°C	30 sec
Annealing	xx°C	30 sec
Extension	72°C	2 min
Final extension	72°C	15 min

(xx - primer specific annealing temperature)

2.4.4 SSCP analysis

SSCP is a method, which has the potential to distinguish single base pair differences between DNA fragments. SSCP is based on the principle that the electrophoretic mobility of a single stranded DNA molecule in a non-denaturing gel is dependent on its size and conformation. The length of a strand, location and number of regions of base pairing determine these secondary and tertiary conformations of the molecule. These conformations are highly dependent on primary sequence. Hence, a mutation at a particular site in the primary sequence can change the conformation of the molecule, which alters its mobility during electrophoresis. (Gasser and Monti, 1997). Generally, the region of interest in the genome or cDNA is amplified by PCR, denatured to form single strands, and analysed by non-denaturing polyacrylamide gel electrophoresis (PAGE, Orita *et al.*, 1989).

RNA was digested with DNaseI, purified by phenol:chloroform extraction and run on an RNA gel (2.4.4). cDNA synthesis was carried out according to Section 2.4.2 followed by PCR (2.3.4) with primers FOR SSCP and REV SSCP. Aquaporin genes *PnAq1-5* were individually subcloned into a pBS vector and used as templates for reference PCR products. Following the PCR, samples were subjected to agarose gel electrophoresis (2.3.1) and the DNA concentration was determined (2.3.1 - 2.3.3). SSCP analysis was carried out on a 10% non-denaturing acrylamide:bisacrylamide gel (28:1) containing 5% glycerol using a Bio-Rad Minigel apparatus (Bio-Rad Laboratories). The gel was pre-run at 60 V for 1 h at 4°C before the samples were loaded. Equal amounts of PCR product were mixed 1:10 with Sample buffer, heated for 10 min at 96°C and chilled on ice. As a marker, 25 ng of the Invitrogen 1 kb DNA Ladder and equal amounts of PCR products of aquaporin genes *PnAq1* and *PnAq4* were mixed and denatured as described above. 5 µL of the samples and the marker mix were loaded onto the gel and electrophoresis was carried out in pre-chilled 1x tris-borate-EDTA (TBE) buffer for 4.5 h at 200 V at 4°C. After electrophoresis, polyacrylamide gels were carefully removed from the glass plate and silver stained as described in Caetano-Anollés and Gresshoff (1994). Following a fixation in 7.5% acetic acid for 20 min, gels were washed with three changes of ddH_2O (5 min each), impregnated with a Silver Nitrate solution for 30 min and rinsed with ddH_2O for 20 sec. Gels were developed in fresh 8°C cold Developer solution until desired band intensity was reached. Finally, the stain was fixed in cold 7.5% acetic acid.

Gels were placed between sheets of cellophane and dried with a Bio-Rad Gel Dryer 583 (Bio-Rad Laboratories) at 75°C for 2 h.

2.4.5 RNA separation by formaldehyde agarose gel electrophoresis

Total RNA was electrophoretically separated under denaturing conditions in a formaldehyde containing agarose gel. Up to 10 µg (11 µL maximum volume) of RNA (2.4.1) was denatured in 9 µL of 12.3 M formaldehyde and 25 µL formamide by heating at 65°C for 15 min according to the procedure described in Sambrook and Russell (2001). The agarose gel was prepared by dissolving 1 g of agarose in 72 mL of sterilised DEPC-treated ddH_2O and cooled to 65°C in a water bath. The following steps were carried out in the fume hood. Once cooled to 65°C, 10 mL of 10x MOPS (3-(N-morpholino)-propanesulphonic acid) buffer was added to the agarose followed by 18 mL of 12.3 M formaldehyde. The gel was poured in the casting apparatus, previously cleaned with 2% hydrogen peroxide to remove RNAses, and allowed to set before pouring 1x MOPS buffer into the casting apparatus to completely cover the gel and the electrodes. RNA gel loading buffer was added to the denatured RNA sample, loaded onto the gel and the gel was electrophoresed at 5 V/cm until the bromophenol dye migrated to two-thirds the length of the gel. The gel was examined under a UV transilluminator and photographed using the GelDoc Imaging system as described in Section 2.3.2.

2.4.6 RNA Blot analysis

RNA blots were prepared by Val Mclean according to protocols described in Shan *et al.* (2004). Plasmid DNA from cDNA clones for use as probes in hybridisation reactions was isolated using the QIAprep Spin Kit miniprep protocol as described in 2.2.1. Inserts were released from cDNA clones by *Eco*RI and *Xho*I double digestion (2.3.6), purified by gel electrophoresis (2.2.5) and labelled with ^{32}P as described in Section 2.2.6.1. The membranes were hybridised according to protocols described in Section 2.3.5.2, and the signals were captured using phosphor-imager screens (2.3.5.3.1). In order to re-use the RNA blots, the hybridised probes were removed completely by submerging the membranes in boiling 0.1% SDS and allowing the solution to cool to RT. Successful removal of labelled probes was confirmed by exposing the washed membranes to phosphor-imaging screens for 24 h and scanning the image.

2.5 Methods - Protein techniques

2.5.1 Protein isolation

P. nicotianae tissue was harvested as described in Section 2.6.1. Vegetative and sporulating hyphal tissue was ground to a fine powder in liquid-N_2 and solubilised in 2D-Rehydration buffer. Proteins of zoospores and germinated cysts were obtained by solubilisation of the frozen pellets in 2D-Rehydration buffer. The suspensions were centrifuged for 10 min at 16,000 x *g* at 4°C to remove insoluble material. The supernatants were stored at –20°C until use. Protein concentrations were determined using the Bradford assay as described in Section 2.5.6.

2.5.2 Peptide design for production of antibodies

Peptides were designed based on hydrophobicity plots, potential post-translation modification sites, antigenicity, required conjugation methods and feasibility of synthesis. Peptides for the production of antibodies were obtained from Sigma-Aldrich.

2.5.3 Peptide conjugation

Peptides were coupled to carrier proteins in order to facilitate a better immune response when injected into antibody producing animals. Conjugation is important as small molecules alone (i.e. peptides) are not very immunogenic and might only trigger a very week immune response. Keyhole limpet hemacyanin (KLH)-coupled peptides were used for immunisation, whereas bovine serum albumine (BSA)-coupled peptides were utilised in ELISA experiments.

2.5.3.1 KLH coupling - MBS method

KLH was coupled to the peptide via the SH group of cysteines present. The method used was based on the m-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS, Pierce, Rockford, IL, USA) coupling method (Figure 2.3, Pierce manual).

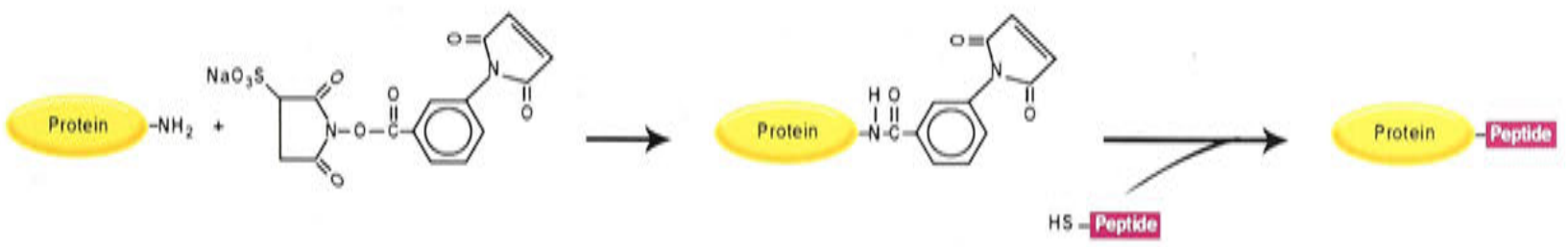


Figure 2.3 MBS coupling (modified from Pierce manual)

To activate the carrier protein, 250 µL sulpho-MBS (2 mg/mL in phosphate buffered saline [PBS]) was added drop-wise to 1 mL KLH (10 mg/mL in ddH_2O). The suspension was incubated for 1 h at RT and mixed occasionally. Activated carrier protein was separated from non-activated protein with a PD10 column (GE Healthcare). To equilibrate, the column was washed with 10 volumes of PBS. The activated carrier protein was loaded onto the column and 1 mL fractions were collected. Fractions from the first peak (monitored with a Pharmacia Ultrospec II at 280 nm) were pooled and stored on ice. 5 mg of peptide was dissolved in 1 mL fresh 100 mM sodium borate, pH 9.3, 100 µL 0.1 M $NaBH_4$ was added and the reaction was incubated at RT for 5 min. In order to inactivate the borohydrate, the pH was reduced to 1 with drops of 1 M HCl and then slowly raised to 7 with 1 M NaOH. The reduced peptide was then coupled to the activated carrier protein. Approximately 2 mL of carrier protein was added to ~1 mL reduced peptide and incubated at RT for 2 h. The suspension was dialysed against three changes of PBS at 4°C to remove uncoupled peptide. The protein concentration was determined according to Section 2.5.6 and coupled peptides were stored in 1 mL aliquots at -20°C.

2.5.3.2 BSA coupling – Glutaraldehyde conjugation

BSA was coupled to the peptide via the NH-group. Glutaraldehyde cross-links primary amino groups of the peptide to those on the carrier. This protocol is based on the Claire Walczak coupling method (http://mitchison.med.harvard.edu/Protocols.htm).

1 mL BSA (2 mg/mL in 0.1 M $NaHCO_3$) was added to 0.5 mL peptide (1 mg/mL in dimethyl sulphoxide [DMSO]) followed by 1 mL of 2% glutaraldehyde in $NaHCO_3$. The mixture was incubated for 90 min at 37°C to allow coupling of the peptide to the carrier protein. In order to block reactive groups, 0.1 volumes of 0.1 M $NaBH_4$ in 0.1 M $NaHCO_3$ was added and the reaction was incubated for 15 min at 37°C. After the incubation time, the same amount of $NaBH_4$ was added. The protein concentration was determined according to Section 2.5.6 and coupled peptides were stored in 0.5 mL aliquots at -20°C.

2.5.4 Production of antibodies

Immunisation, fusion of spleen cells, hybridoma cell culturing and sera collection was conducted by Ms Janet Elliott (Plant Cell Biology Group, RSBS, ANU). All animal work was approved by the ANU Animal Experimentation Ethics Committee .

2.5.4.1 Polyclonal antibodies

Polyclonal antibodies were raised in rabbits against peptides Pep-PP1, Pep-PP2, Pep-Aq1 and Pep-Aq2. Two animals were used for each PnVPP peptide, aquaporin peptides were injected together into two animals to increase antigenicity and the likelihood of obtaining a well performing antibody.

On the first day, a bleed was taken to obtain a pre-immune serum and each rabbit was injected subcutaneously with 400 µg of KLH-coupled peptide (2.5.3.1) emulsified in an equal volume of Freund's complete adjuvant. Subsequent boost injections were performed every 3 weeks with KLH-coupled peptide and Freund's incomplete adjuvant. Test bleeds were taken from the ear vein 7 days after each injection and analysed following the ELISA protocol described in Section 2.5.7. For larger bleeds (~25 mL) rabbits were tranquilised with acepromazine and blood was taken from the central ear artery. The terminal bleed was carried out by heart puncture after anaesthesia of the rabbit with isoflurane.

2.5.4.2 Monoclonal antibodies

In addition to polyclonal antibodies, it was also attempted to produce monoclonal antibodies from peptides Pep-PP1 and Pep-PP2. Immunisation of two Balb/c mice per peptide was carried out according to the schedule described above with some modifications. For boosts, 50 µg of coupled peptide was injected intraperitoneally and bleeds were taken from the orbital route. Bleeds were analysed by ELISA and immunofluorescence microscopy of zoospores (2.5.7 and 2.5.8). The spleen of positive mice was removed and fused with murine myeloma cells (Sp_2O) according to standard protocols. Fused cells were plated in a microtiter plate and checked for hybridoma cell growth between days 5 and 10 after fusion. Screening of growing hybridoma cells was carried out by immunofluorescence microscopy of *P. nicotianae* zoospores as described in Section 2.5.8. Hybridomas producing the desired antibody were cloned, re-screened and finally grown in large scale cultures.

2.5.5 Purification of polyclonal antibodies

Sera of rabbits were centrifuged at 3000 x *g* for 30 min at 4°C to remove lipids and cell debris. The supernatant was transferred to a beaker, gently stirred at 4°C and ammonium sulphate was added slowly until a concentration of 45% was reached. The suspension was stirred for 1 h at 4°C to allow complete precipitation and then centrifuged for

20 min at 10,000 x *g* at 4°C. The supernatant was discarded and the pellet resuspended in PBS at 0.25 volumes of the starting volume. The solution was dialysed against three changes of PBS overnight. The following day, the suspension was filtered twice through Millipore filters (5 µm and 0.8 µm, Billerica, MA, USA) and finally purified on a HiTrap Protein G HP column (GE Healthcare). The column was equilibrated with 10 mL PBS and the antibody solution was loaded onto the column. In order to remove unbound material and impurities, the column was washed with 5 mL PBS. Antibodies were eluted by addition of 5 mL Elution buffer. 1 mL fractions were collected into microfuge tubes containing 50 µL 1 M Tris-HCl, pH 9 to neutralise. The fractions were tested by Bradford assay (2.5.6) and SDS-PAGE (2.5.10). Fractions were pooled and the final protein concentration was determined. Antibodies were stored in 1 mL aliquots at -20°C.

2.5.6 Bradford assay

Protein concentrations were determined using the Bio-Rad Bradford Protein Assay (Bio-Rad Laboratories) reagent with BSA dilutions as standard (Bradford, 1976).

The Bradford reagent was diluted 1:5, filtered through Whatman #2 filter paper and stored on ice. For the production of a standard curve, 1 mg/mL BSA was made up in PBS and diluted according to Table 2.2.

Table 2.2 BSA dilutions for standard curve

Final BSA µL	Volume of 1 mg/mL BSA	Volume of 50 µg/mL BSA	PBS µL
0			100
5		60	540
10		120	480
25		300	300
50	30		570
75	45		555
100	60		540
200	120		480

Protein samples were diluted as required (generally, 1:50 – 1:200) and kept on ice. 100 µL of BSA standards and diluted proteins were transferred to a microtiter plate in triplicates. 100 µL of the Bradford reagent was added to each well, reactions were mixed carefully and the absorbance at 595 nm was measured using a Bio-Rad Microplate Reader (Model 680, Bio-Rad Laboratories). Protein concentrations of samples were calculated using MS Excel.

Protein concentrations of samples containing 2D-Rehydration buffer were determined using a modified Bradford assay with BSA diluted in 2D-Rehydration buffer as a standard and with the addition of HCl to compensate for the presence of detergent (modifications based on Bio-Rad handbook).

2.5.7 ELISA

ELISAs were used to monitor the antibody response of rabbits and mice towards the injected peptide. Wells of a 96-well microtiter plate were coated with 50 µL of 10 µg/mL BSA-coupled peptide in PBS (2.5.3.2) and incubated at RT for 1 h. The unbound peptide was removed by washing the plate with three changes of PBS. Antigen solution and washes were discarded by inverting the plate and tapping the bottom on paper towels to remove any drops. Starting from a 1:100 dilution of the sera (in PBS plus Tween20 [PBST]) serial dilutions were prepared in a separate microtiter plate by diluting 1:2 each row, thus yielding dilutions from 1:100 – 1:204,800. 50 µL of the prepared dilutions were transferred to the antigen coated wells and the plate was incubated for 1 h at RT. Following three washes in PBST, the wells were incubated for 1 h at RT with 50 µL of either anti-mouse or anti-rabbit biotin conjugated antibody diluted 1:300 in PBST. After the incubation period plates were washed with three changes of PBST and a streptavidin-horse radish peroxidase tag (HRPO) was added to the biotinylated antibodies by incubating with 50 µL of 1:5000 strepatvidin-biotin-HRPO (in PBST) for 30 min at RT. The plate was washed another three times with PBST and the colour reaction was started by addition of 100 µL 2,2'-azino-di-(3 ethylbenzthiazoline sulphonic acid (ABTS) substrate. The reaction was stopped after 20 min by addition of 50 mM NaF and the absorbance was read at 405 nm using the Bio-Rad Microplare Reader Model 680. Data was analysed using MS Excel.

For calculation of the antibody titer, absorbance readings were plotted vs. the antiserum dilution and the inflection point of the graph was estimated. The titer was interpolated by drawing a line down to the X-axis (Perkin-Elmer manual).

2.5.8 Immunofluorescence microscopy

Indirect immunofluorescence labelling was carried out as described in Hardham *et al.* (1991) with some modifications. *P. nicotianae* zoospores were fixed in 4% formaldehyde or in 4% formaldehyde combined with 0.2% glutaraldehyde in 50 mM piperazine-1,4-

bis(2-ethanesulphonic acid) (PIPES) buffer pH 7.0 for 30 min at RT. The cells were collected by centrifugation at 1000 x *g* for 2 min and the cell pellets washed once in 100 mM PIPES pH 7.0 and twice in PBS. The final samples were resuspended in ddH_2O and 10 µL aliquots were applied to each well of Multitest slides (ICN Biomedicals, Irvine, CA, USA). Samples were allowed to air dry for approximately 1 h at 37°C, cells were rehydrated in PBS and incubated in 15 µL primary antibody in a moist chamber at 37°C for 1 h. Depending on application sera were diluted in PBS containing 1% BSA and 0.1% gelatine (PBSBG). After two rinses (2 min, 5 min) in PBS, 15 µL of secondary antibody (either either sheep F(ab') anti-mouse or anti-rabbit antibody conjugated to fluorescein isothiocyanate [FITC], SAM-FITC, SARa-FITC; Chemicon, Temecula, CA, USA) diluted 1:30 in PBSBG were added and incubated at 37°C for 1 h. The slides were rinsed twice in PBS (2 min, 5 min) and cells were DAPI-stained by incubation in a 0.1 µg/mL DAPI solution for 2-3 min. The cells were washed once in PBS and ddH_2O and then mounted in a glycerol-based mounting medium containing 0.1% of the anti-fade agent ρ-phenylenediamine. The slides were left to set at 4°C in the dark. Cells were viewed using a Zeiss Axioplan microscope (Zeiss, Wetzlar, Germany) equipped with differential interference contrast (DIC) and bright field optics, with a 50 W mercury vapor lamp and incident-light fluorescence illumination and an FITC filter set. Digital images were collected using a Princeton Instruments MicroMax camera (Tucson, AZ, USA) and the MetaMorph software (Molecular Devices Corp., Sunnyvale, CA, USA).

2.5.9 Immunodot blot

BSA-coupled peptides (2.5.3.2) and zoospore protein extract (2.6.1.3) was used for immunodot blot experiments. A Hybond-C nitrocellulose membrane (GE Healthcare) was cut to size, equilibrated in tris buffered saline (TBS) and fixed in the dot blot apparatus (Bio-Rad Laboratories). Wells were washed with TBS and then drained via a vacuum pump. 10 µg of protein or peptide in TBS were applied to the wells and allowed to settle for 30 min at RT. Wells were drained and the membrane was taken out of the apparatus, marked and cut into strips and left to dry at RT. Primary antibodies were pre-absorbed with peptides according to Section 2.5.11 and the membranes were labelled and developed as described in Section 2.5.10.

2.5.10 SDS-PAGE and western immunoblot

P. nicotianae tissues were harvested as described in Section 2.6.1 and protein concentrations of samples were estimated by Bradford assay as described in Section 2.5.6.

Samples were diluted in an equal volume of 2x Sample buffer boiled for 3 min and chilled on ice. Samples were loaded onto SDS-PAGE gels (10 – 15% depending on application) in a Bio-Rad Minigel apparatus and run for 1-3 h at 100 V (according to Laemmli, 1970). Proteins were transferred onto a Hybond C nitrocellulase membrane either in 1.5 h at 100 V or overnight at 25 V with the current limited to 0.4 mA. Protein molecular weight markers were used as an indicator for protein transfer from the gel to the membrane. Non-specific binding was blocked by incubating the membrane in TBST (TBS plus Tween20) containing 5% skim milk powder for 2 h at RT. The membranes were washed three times (5 min each) in TBST and cut into strips. One strip was stained in 0.2% Coomassie brilliant blue in a 0.5% solution of acetic acid in 30% ethanol for 1 min then destained in 0.5% acetic acid in 30% ethanol for 1 h. Remaining strips were incubated individually for 1 h at RT in the primary antibody with varying dilutions (in 1% BSA in TBS) on clean Nescofilm (Azwell Inc., Ozaka, Japan) in a humid chamber. Following three washes (5 min each) in TBST, the strips were incubated for 1 h at RT in either SAM or SARa antibody conjugated to AP (SAM-AP, SARa-AP; Chemicon) diluted 1:10,000 in TBST. Membrane strips were washed once in TBST and once in TBS (5 min each) and then developed by adding the AP substrate (one FastRed® tablet per 10 mL ddH_2O). The colour reaction was stopped by removing the developing solution and rinsing the membranes in two changes of ddH_2O. The membranes were dried between filter paper.

2.5.11 Peptide competition assay

Peptide competition assays were performed in order to demonstrate the specificity of antibodies towards their peptides. Antibodies were used at sub-saturating concentrations and incubated with a range of peptide concentrations (e.g. 50- and 200-fold molar excess). The required amount of peptide was calculated according to the following formula:

$$Amount\ of\ \ peptide\ required\ \mu g = \frac{\mu g\ \ Ab * 2(binding\ sites) * x(excess\ \ peptide) * 110 MW\ \ of\ \ an\ aa * y(\#aa\ \ in\ \ peptide)}{150000(MW\ \ of\ \ Ab)}$$

Antibodies were diluted in PBSBG to yield the required final antibody concentration and the calculated amount of peptide was added. Reactions were incubated at RT for 2 h with gentle rocking and then centrifuged at 16,000 x *g* for 15 min at 4°C to remove any antibody/protein complex formed during the incubation period. This pre-absorbed antibody was used in immunodot blots and western immunoblots (2.5.9 and 2.5.10, respectively) as primary antibody.

2.6 Methods - *Phytophthora* and Bacteria

2.6.1 *Phytophthora* strain and culture conditions

P. nicotianae strain H1111 (ATCC MYA 141) used in this study was originally isolated from tobacco plants in Queensland, Australia, and kindly provided by Prof David Guest (University of Sydney, Australia). The working stock cultures were maintained on V8 agar at 25°C in the dark and subcultured every 7 days. Long-term stock cultures were stored at 15°C on V8 agar slopes under paraffin oil.

P. sojae strain 1176 was isolated from soybean and supplied by Dr Brett Tyler (Virginia Bioinformatics Institute [VBI], Blacksburg, VA, USA). Stock cultures were maintained as described for *P. nicotianae*.

2.6.1.1 Vegetative hyphae

In order to obtain vegetative hyphae of *P. nicotianae*, 7 x 1 mm^2 agar plugs from freshly grown stock cultures were incubated on discs of miracloth on V8 agar in 90 mm polystyrene Petri dishes for 3 days at 25°C in the dark. The miracloth discs were transferred individually into 250 mL flasks containing 100 mL V8 broth and grown with shaking for 5 days at 23°C in the light. The medium was changed every 24 h until harvesting. Tissue representing vegetative hyphae was harvested after this incubation period by removing the agar plugs and scraping the hyphae tissue off the miracloth disc. Excess medium from the hyphal tissue was removed using vacuum filtration and the tissue was snap-frozen in liquid-N_2 and stored at –80°C.

2.6.1.2 Sporulating hyphae

In preparation for the induction of sporulation of *P. nicotianae* hyphae, cultures were prepared as for vegetative hyphae and incubated for 5 days in the dark at 25°C. The miracloth discs were transferred individually into 90 mm Petri dishes containing

approximately 30 mL V8 broth and grown in still cultures for 24 h at 23°C in the light. The presence of sporangia was verified microscopically and cultures were harvested as for vegetative hyphae.

2.6.1.3 Zoospores

Zoospores were produced on mycelium-containing miracloth discs after incubation in V8 broth under light at 23°C for 7–21 days. Zoospores were released by rinsing discs three times in cold sterile water and incubation in 10 mL cold water in the light at 15°C for at least 90 min. Zoospores in solution were combined in 50 mL screw-cap Falcon tubes and the zoospores concentrated by centrifugation for 1 min at 1800 x *g* at RT. The supernatant was decanted and the pellet frozen in liquid-N_2. For the use in immunofluorescence microscopy zoospores were fixed as described in Section 2.5.8.

2.6.1.4 Germinated cysts

For production of germinated cysts, V8 broth was added to a zoospore suspension in a ratio of 50:50 and tubes were vortexed 5x for 10 sec to induce encystment of the zoospores. Encystment was monitored using a Zeiss Axioplan microscope. Cysts were allowed to germinate for 2 h at RT and were harvested by centrifugation at 2500 x *g* for 1 min at RT and stored at –80°C after snap-freezing in liquid-N_2.

2.6.2 Microsomal preparation of *Phytophthora* zoospores

Zoospores of approximately 20 plates were pooled and LiCl was added to a final concentration of 100 mM. Cells were centrifuged immediately at 750 x *g* for 5 min at 4°C. The pellet was resuspended in ~500 µl of 10x Osmoticum buffer and sonicated for 40 sec (8 x 5 sec bursts) to break the plasma membrane. Lysed cells were transferred to pre-cooled centrifuge tubes and centrifuged at 6000 x *g* for 15 min at 4°C. The supernatant was re-centrifuged at 85,000 x *g* for 1 h at 4°C, the resultant microsomal pellet was resuspended in Resuspension buffer and stored at -80°C until use.

2.6.2.1 Determination of pyrophosphatase activity

Activity of the V-PPase in *Phytophthora* zoospore microsomal fractions (2.6.2) was determined using the EnzCheck™ Phosphate Assay Kit (Molecular Probes, Oregon, USA). The assay is based on the measurement of free phosphate (P_i) which is released upon pyrophosphatase activity (Figure 2.4 a). In this assay, 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG) is converted to ribose-1-phosphate and 2-amino-

6-mercapto-7-methylpurine by the enzyme purine nucleoside phosphorylase (PNP) in the presence of P_i (Figure 2.4 b). The absorption maximum is shifted from 330 nm for the substrate to 355 nm for the product as a result of the reaction. Kinetic assays were performed by addition of inorganic phosphate (PP_i) and the experimental enzyme according to manufacturer's instructions.

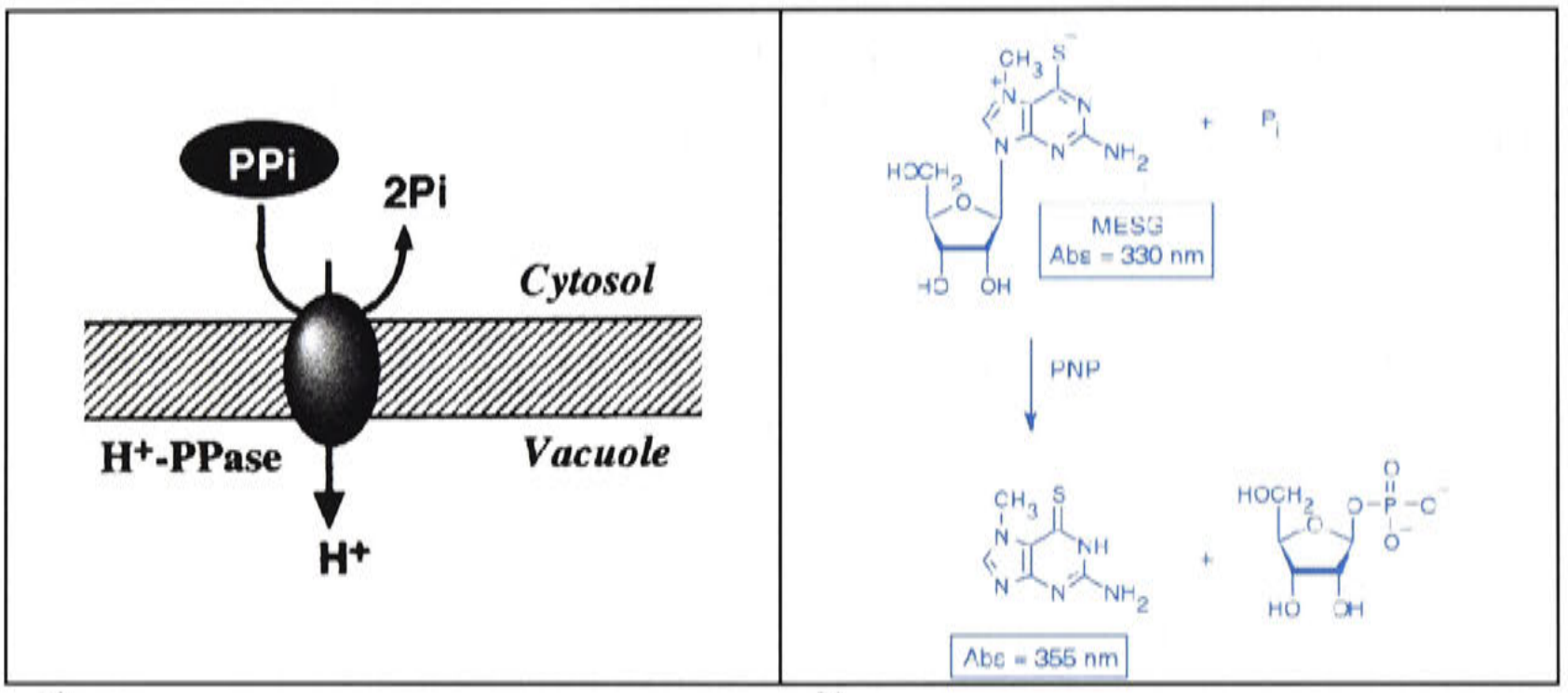


Figure 2.4 Pyrophosphatase assay (a) Production of P_i by V-PPase activity (Source: Maeshima, 2000), (b) conversion of MESG to ribose-1-phosphate and 2-amino-6-mercapto-7-methylpurine by PNP in the presence of P_i.(Source: Molecular Probes, 2002)

Microsomal zoospore fractions were prepared according to Section 2.6.2 and the protein concentration was determined by Bradford assay (2.5.6). Reagents supplied with the kit were prepared as follows. A 5x MESG stock solution (1 mM) was made up and 1 mL aliquots were stored at -20°C. Aliquots were thawed as required immediately prior to use. Lyophilized PNP enzyme was dissolved in ddH_2O and stored at 4°C (0.1 U/µL).

A standard curve for P_i was generated using the supplied phosphate standard. The 50 mM phosphate stock was diluted with ddH_2O to produce a standard curve ranging from 0 to 150 µM. Standards were made up to a volume of 70 µL and dispensed into a microtiter plate in replicates of three. A master mix containing 1x Reaction buffer, 0.2 mM MESG and 0.2 U PNP was prepared in a volume of 30 µL and the reaction was started by addition of the substrate mix to the standard dilutions. Reactions were incubated for 30 min at RT after which their endpoint absorbance at 360 nm was determined using a µQuant Scanning Microplate Spectrophotometer (Biotek Instruments, Winooski, VT, USA). The activity of microsomal samples was measured by

the addition of PP_i solution (3.5 mM). Approximately 5 µg of protein was used per reaction and combined with 30 µL of freshly prepared master mix. The volume was made up to 80 µL with ddH_2O and the reaction was pre-incubated for 10 min at RT to eliminate potential P_i contamination. The assay was initiated by addition of 20 µL PP_i substrate and the microtiter plate was immediately placed into the µQuant Scanning Microplate Spectrophotometer were kinetic readings in 2.5 min intervals were performed at 360 nm. Data capture and reader control was carried out using KC4 software (Bio-Tek Instruments). Activity inhibition tests were performed in a similar manner. Several dilutions of the test substance were added to the protein and the volume was adjusted to 80 µL. Reactions were pre-incubated for 10 min to allow for P_i contamination removal and reaction between the V-PPase and the test substance. The reaction was initiated by the addition of 20 µL PP_i (300 µM) substrate and kinetic readings were performed as above.

Analysis of data obtained was carried out in MS Excel. For the standard curve, the background absorbance determined by a no-phosphate control was subtracted from each sample and data was plotted to construct a standard curve (PP_i concentration vs. A_{360nm}). Data of microsomal preparations and inhibition test were plotted (A_{360nm} vs. time) after subtracting values of no-substrate controls.

2.6.3 Video analysis of *Phytophthora* zoospores

In vivo video analysis was carried out using *P. nicotianae* zoospores in order to test inhibitor substances prior analysed in *in vitro* experiments. *P. nicotianae* was cultured as described in Section 2.6.1. Miracloth discs were prepared for zoospore release by rinsing three times in cold sterile ddH_2O, however, plates were not placed in the light, thus slowing down the release process. A small tuft of hyphae was placed on a microscope slide and left on ice for several minutes to facilitate zoospore release. Zoospores were either left untreated or were exposed to potential inhibitor substances. Sporangia and zoospores were observed with DIC optics on a Zeiss Axioplan microscope. Zoospores trapped between hyphae were located at a 100x magnification and video recordings were made with a Sony DXC-327P camera. Video material was analysed with Adobe Premiere 6.5 by marking frames that corresponded to the contraction of the WEV and the water expulsion cycle times were calculated.

2.6.4 Competent bacteria and transformation

High efficiency transformation (~100 transformants per µg of circular plasmid DNA) of *E. coli* strains was achieved by electroporation using a Bio-Rad Gene Pulser electroporator with a Bio-Rad Pulse Controller (Bio-Rad Laboratories). 1-2 µL DNA was mixed with a 40 µL aliquot of electrocompetent DH10β *E. coli* cells (see Sambrook and Russell, 2001) and left on ice for 1 min. The mixture of cells and DNA was transferred to a cold Bio-Rad electroporator cuvette and subjected to an electrical pulse field strength of 9 kV/cm (the Gene Pulser was set to 25 µF and 1.8 kV, and the Pulse Controller was set at 200 Ω, using a cuvette of 1 mm gap width). 1 mL of SOC media was immediately added to the cells, the mixture was transferred to a 1.5 mL microfuge tube and incubated at 37°C for 1 h with moderate shaking to allow expression of the antibiotic resistance gene and recovery of antibiotic containing transformed *E. coli* cells. An aliquot of the transformation was plated on selective medium agar plates. To facilitate blue/white screening of transformants, 5 µL of 1 M isopropyl-β-D-thiogalactoside (IPTG) and 20 µL (50 µg/mL) 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) were spread on the agar plates 30 min prior to use. After spreading the transformed cells, the agar plates were incubated at 37°C overnight. White colonies indicating presence of inserts were selected for plasmid DNA isolation (2.2.1).

2.6.5 Storage of *E. coli* cultures

For short-term storage, plates with *E. coli* colonies were stored at 4°C, usually up to 3 months, after which they were autoclaved and discarded. For long-term storage a culture was prepared from an inoculum of a colony forming unit grown overnight in LB broth plus the appropriate antibiotic at 37°C and vigorous shaking at 200 rpm. The next day, a 600 µL aliquot of culture was suspended in 400 µL 50% glycerol and stored at -80°C.

2.7 Bioinformatic analysis

2.7.1 Design of custom primers and DNA sequencing

Custom primers were designed using the programs Oligo Explorer 1.1.0 and Oligo Analyzer 1.0.2 (Teemu Kuulasmaa, Finland) and obtained from Sigma-Genosys (Castle Hill, NSW, Australia). Plasmid DNA at a concentration of 200 ng/µL and primers at a concentration of 0.8 pmol/µL were sequenced at the Australian Genome Research

Organization Sequencing Facility (University of Queensland, Brisbane, Australia). For a list of primers used in this study see Appendix E.

2.7.2 Sequence data analysis

Nucleotide sequences obtained were translated into amino acid sequences using the program BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html; Hall, 1999). DNA, protein, and conserved domain database (CDD) searches were carried out by the BLAST programs through the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/BLAST; Altschul *et al.*, 1997). The molecular weight and theoretical isoelectric point (pI) of proteins was calculated with ProtParam on the ExPasy server (http://ca.expasy.org/tools/protparam.html; Wilkins *et al.*, 1999). Virtual cloning and graphical representation of vectors was performed by using pDraw32 (AcaClone Software; http://www.acaclone.com/).

The putative membrane topology and secondary structure of proteins was modelled using TopPred II version 1.3 (http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html; Claros and von Heijne, 1994), SOSUI (http://sosui.proteome.bio.tuat.ac.jp/sosui_submit.html; Hirokawa *et al.*, 1998), SMART (http://smart.embl-heidelberg.de/; Schultz *et al.*, 2000; Letunic *et al.*, 2004), PHD (http://www.embl-heidelberg.de/predictprotein/predictprotein.html; Rost *et al.*, 1995), HMMTOP (http://www.enzim.hu/hmmtop/; Tusnady and Simon, 1998), SPLIT (http://split.pmfst.hr/split/; Juretic *et al.*, 2002), PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/; McGuffin *et al.*, 2000; Jones, 1999), HMMER (http://hmmer.wustl.edu/; Eddy, 1998), DAS (http://www.sbc.su.se/~miklos/DAS/; Cserzo *et al.*, 1997), TMpred (http://www.ch.embnet.org/software/TMPRED_form.html; Hofmann and Stoffel, 1993) and PRED-TMR (http://o2.biol.uoa.gr/PRED-TMR/; Pasquier *et al.*, 1999). The program CoPreTHi (http://o2.biol.uoa.gr/CoPreTHi/Main.html; Promponas *et al.*, 1999), a web application which combines the results of methods that predict the location of transmembane segments in protein sequences into a joint prediction histogram, was used with selected prediction output files.

2.7.3 Database analysis

Other *Phytophthora* genomes were analysed for representation of V-PPase and aquaporin genes using the *P. sojae* (http://genome.jgi-psf.org/sojae1/sojae1.home.html)

and *P. ramorum* (http://genome.jgi-psf.org/ramorum1/ramorum1.home.html) databases. *P. infestans* EST contigs were obtained by mining the *Phytophthora* Functional Genomics Database (PFGD) at https://www.pfgd.org, which comprises ESTs and genomic sequences from *P. infestans* and *P. sojae* (Kamoun *et al.*, 1999; Waugh *et al.*, 2000; Kamoun, 2003). Furthermore, the genomic database of the diatom *Thalassiosira pseudonana* (http://genome.jgi-psf.org/thaps1/thaps1.home.html) was analysed for the existence of aquaporin and V-PPase genes.

2.7.4 Phylogenetic analysis

Homologues of V-PPase and aquaporin genes from other organisms used in phylogenetic analysis (2.7.4) were retrieved from GenBank (http://www.ncbi.nlm.nih.gov). Multiple sequence alignments were performed by ClustalX 1.83 (Thompson *et al.*, 1997) and for phylogenetic tree reconstruction programs of the PHYLIP package v. 3.5 (http://evolution.genetics.washington.edu/phylip.html, Felsenstein, 1989) was used. Phylogenetic analysis was performed by the Neighbor-Joining (NJ) method. Gaps in the multiple sequence alignments were trimmed with the program parseClustal (courtesy of Allan Dickerman, VBI, Blacksburg, VA, USA) to a sequence density of >50% per column. The resulting alignment was used to generate a NJ phylogram with 1,000 bootstrap replicates. For details see Appendix H.

Chapter 3

Vacuolar H^+-translocating pyrophosphatase in *P. nicotianae*

Chapter 3: Vacuolar H^+-translocating pyrophosphatase in *P. nicotianae*

3.1 Introduction

3.1.1 Transport across cell membranes

Plasma membranes and organelle membranes act as barriers between the inside of the cell and its surroundings or organelle contents. They are highly selective permeable membranes with regards to which molecules are allowed to cross and in which direction they are allowed to travel. A great number of molecules must transfer between the inside and outside of a cell at any given time, to maintain electrochemical and proton gradients across membranes necessary for homeostasis of the cell and organism (Lodish, 2000).

Transport across cell membranes occurs in two ways – active or passive (Figure 3.1). Passive transport, like the process of diffusion, does not require any energy from the cell as molecules move down a concentration or electrochemical gradient. Active transport, however, requires the cell to spend energy, often in the form of ATP, as ions or molecules are moved against a concentration gradient. Plasma membranes contain a number of distinct and highly selective carriers, ion channels and ion pumps, which contribute to the redistribution of charge and ion concentrations between compartments within the cell and between the cell and its environment (McIntosh *et al.*, 2001). Proton pumps are of particular interest as they establish and maintain ionic and electrochemical gradients in cells and their compartments. As a result, the ionic composition of the cytosol usually differs greatly from that of the surrounding medium (Lodish, 2000). In animal cells, the primary source for the generation of those gradients is ATP used by an array of ATPases (P-type: ubiquitous, phosphorylated subunit, F-type: thylakoid and mitochondrial membranes, and V-type: vacuolar membrane; EC 3.6.1.3). In plants, fungi and bacteria, however, the H^+-gradient is established by not only the ATPases but also by a vacuolar-type inorganic pyrophosphatase (V-PPase; EC 3.6.1.1) which utilises exclusively inorganic pyrophosphate (PP_i) as the energy source. Both enzymes catalyse electrogenic H^+-translocation from the cytosol to the vacuole lumen to generate an inside-acid pH difference (ΔpH) and an inside-positive electrical potential difference ($\Delta\Psi$) which is employed to energize a wide range of secondary, $\Delta\Psi$- and/or ΔpH-coupled, transport processes (Rea and Poole, 1993, Figure 3.2).

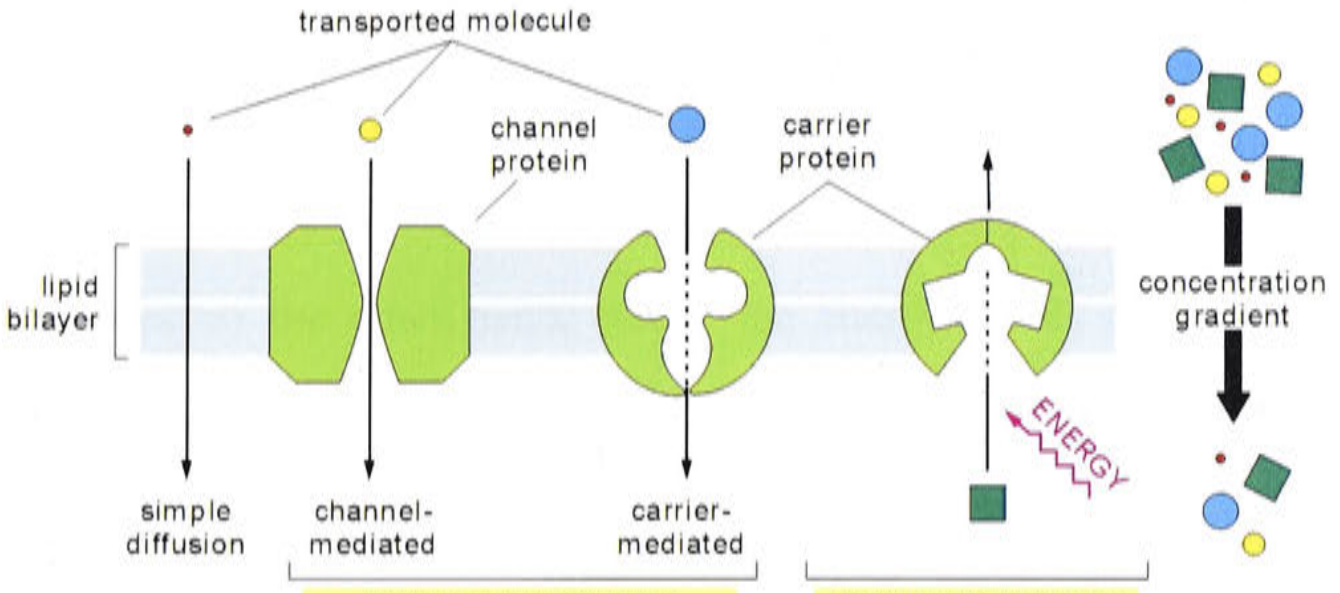


Figure 3.1 Comparison of passive and active transport across membranes. If uncharged solutes are small enough, they can move down their concentration gradients directly across the lipid bilayer itself by simple diffusion. Most solutes, however, can cross the membrane only if there is a membrane transport protein (a carrier protein or a channel protein) to transfer them. As indicated, passive transport, in the same direction as a concentration gradient, occurs spontaneously, whereas transport against a concentration gradient (active transport) requires an input of energy. Only carrier proteins can carry out active transport, but both carrier proteins and channel proteins can carry out passive transport. (Source: http://www.accessexcellence.org/RC/VL/GG/passActiveTrans.html)

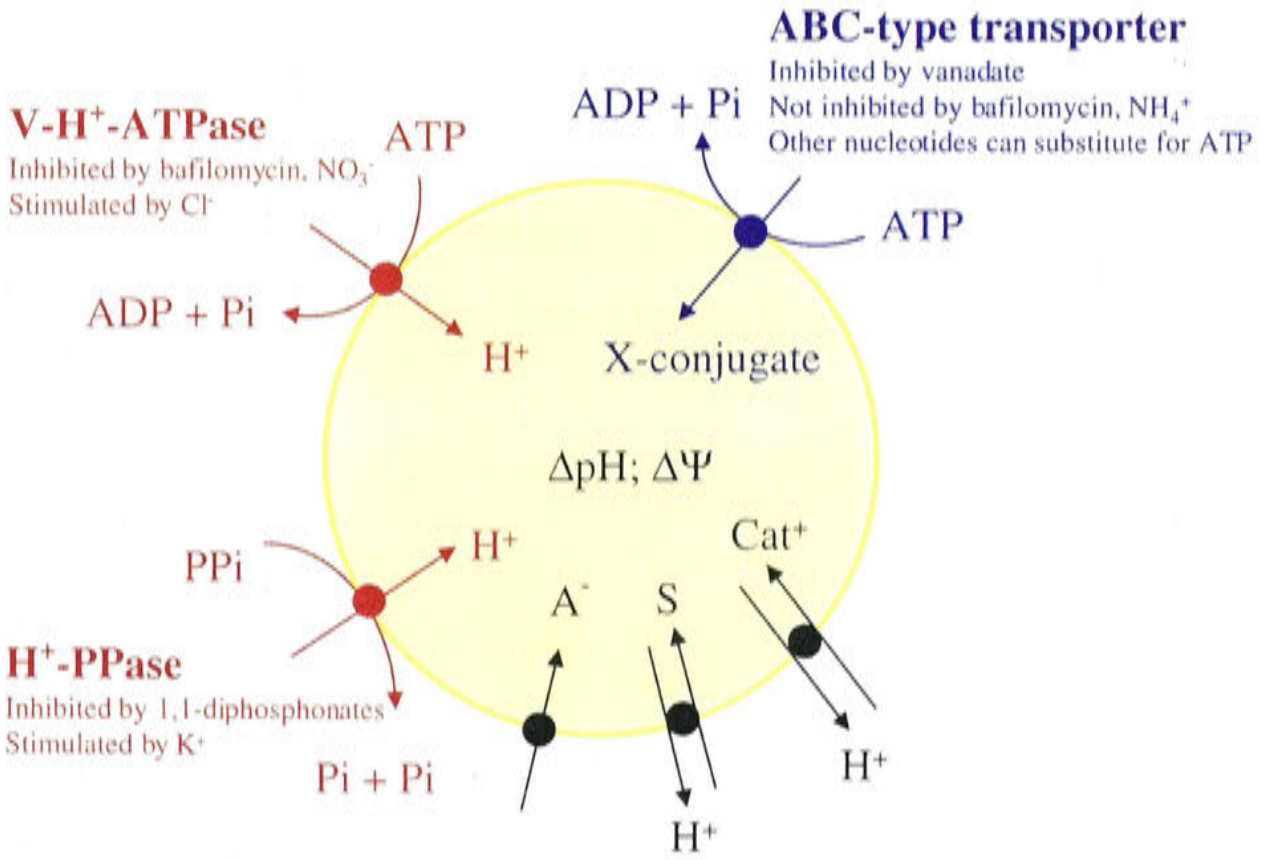


Figure 3.2 Overview of transport processes. Proton pumps establish an electrochemical gradient (red), secondary energized uptake mechanisms (black), and directly energized, ABC-type transporters (blue). S, neutral solute; A^-, anion; cat^+, cation; X-conjugate, conjugate of a compound X (secondary metabolite or xenobiotic) with a hydrophilic compound such as glucose, glutathione, an amino acid, malonate, or sulphate. (Source: Martinoia *et al.*, 2000)

3.1.2 Pyrophosphatases

The class of phosphoryl transfer enzymes is one of the largest groups of enzymes and includes soluble inorganic pyrophosphatase (sPPase) as well as V-PPases. sPPases are ubiquitous proteins found in all organisms from bacteria to humans, which utilise PP_i and release the energy generated by cleavage of the high energy phosphoanhydride bond as heat (Cooperman *et al.*, 1992; Maeshima, 2000; Figure 3.3). Membrane bound pyrophosphatases on the other hand, couple the energy of PP_i hydrolysis to proton movement across biological membranes (Baltscheffsky *et al.*, 1999; Perez-Castineira *et al.*, 2002).

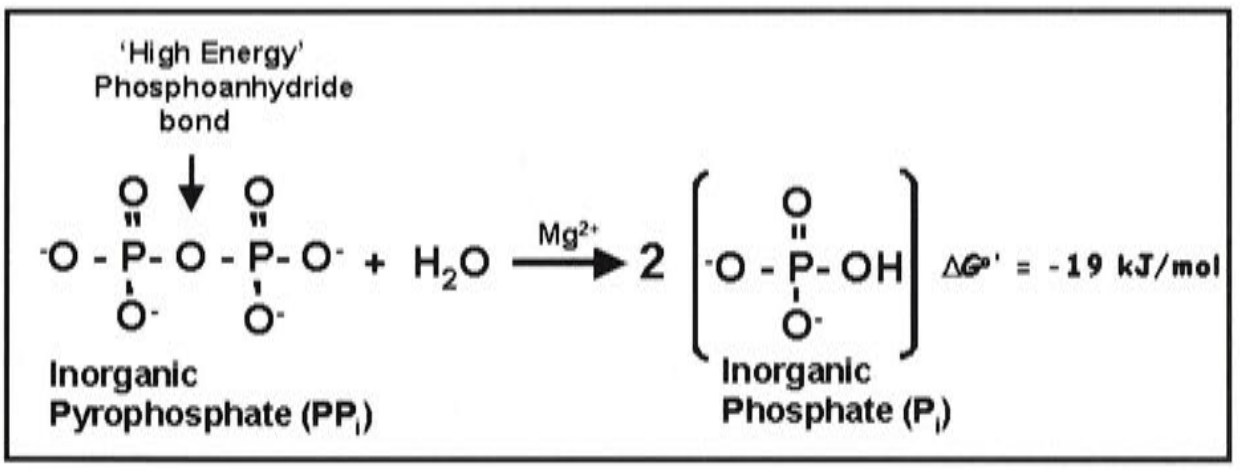


Figure 3.3 Pyrophosphate hydrolysis is catalysed by inorganic pyrophosphatase, a highly exergonic reaction. (Source: http://www.plantphys.net/article.php?ch=e&id=124)

The process of membrane energising and/or acidification of cell compartments to maintain the concentration gradients of Na^+, K^+, H^+, and Ca^{2+} requires a significant amount of energy expenditure. It has been shown, that erythrocytes consume up to 50% of the available intracellular ATP for ion transport, and active nerve cells even require as much as 75% (Lodish, 2000; Alberts *et al.*, 2002). Thus, the ability to utilise PP_i as an alternative energy source may ensure the maintenance of vital functions when ATP demand for other cellular activities is high (McIntosh and Vaidya, 2002). As PP_i is a by-product of a range of major biosynthetic pathways, it is not only readily available but it also serves as an inexpensive, abundant source of energy compared to ATP. Metabolic pathways such as DNA and RNA synthesis, the aminoacetylation of tRNAs, activation of fatty acids by thiokinases to form CoA esters and furthermore synthesis of sucrose, cellulose and starch, produce copious amounts of this energy source (Maeshima, 2000; McIntosh and Vaidya, 2002; Figure 3.4, Figure 3.5).

amino acid activation (aminoacyl-tRNA synthetase)
amino acid + t RNA + ATP ⟶ amino acyl-tRNA + AMP + PPi

RNA synthesis (RNA polymerase)
$(NMP)_n$ + NTP ⟶ $(NMP)_{n+1}$ + PPi

DNA synthesis (DNA polymerase)
$(dNMP)_n$ + dNTP ⟶ $(dNMP)_{n+1}$ + PPi

ADP-glucose formation (ADP glucose pyrophosphorylase, starch synthesis)
glucose-1-phosphate + ATP ⟶ ADP-glucose + PPi

UDP-glucose formation
(UDP glucose pyrophosphorylase, sysntheses of cellulose and sucrose)
glucose-1-phosphate + UTP ⟶ UDP-glucose + PPi

formation of fatty acyl-CoA (fatty acyl-CoA synthetase)
fatty acid + CoA + ATP ⟶ fatty acyl-CoA + AMP + PPi

Figure 3.4 Main steps of PP_i production in plant cells. (Source: Maeshima, 2000)

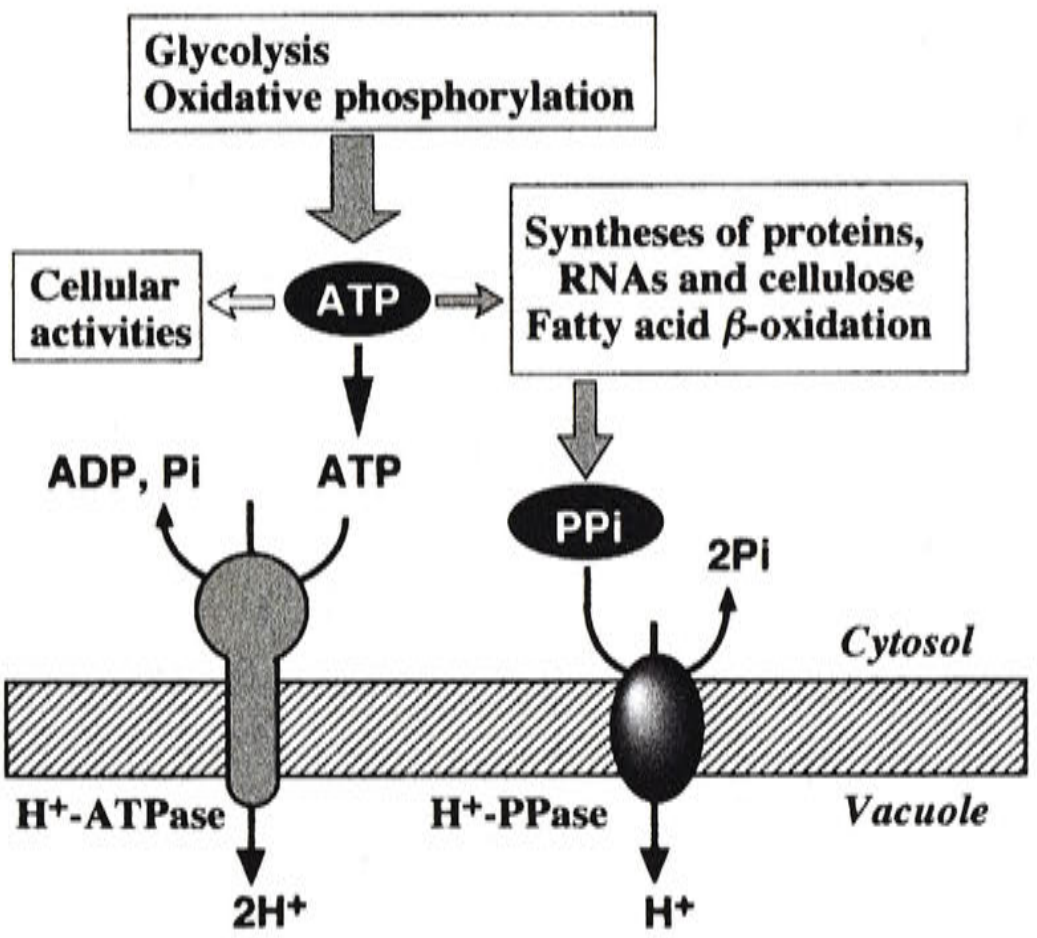


Figure 3.5 PP_i production and vacuolar membrane H^+-PPase and H^+-ATPase. PP_i is supplied as a by-product of biosyntheses of macromolecules such as RNAs, proteins, and cellulose and L-oxidation of fatty acids. (Source: Maeshima, 2000)

3.1.3 Cloning of V-PPases and their distribution in organisms

sPPases have been shown to form a family of homologous enzymes and have been reported in the cytosol of animal and fungal cells, in mitochondria of diverse eukaryotes and in prokaryotes and plastids (Maeshima, 2000; Perez-Castineira *et al.*, 2002). Members of a second group of pyrophosphatases, termed 'membrane associated PPases', closely resembling the sPPase in size, sequence and hydrophobicity (McIntosh and Vaidya, 2002), have been isolated from yeast (Mansurova, 1989) and pea mitochondria (Zancani *et al.*, 1995), spinach thylakoid membranes (Jiang *et al.*, 1997) and rat liver Golgi apparatus (Brightman *et al.*, 1992). However, these PPases did not show any proton translocation activity and it has been postulated that they form complexes with other integral membrane proteins leading to their association with membranes (Maeshima, 2000; McIntosh and Vaidya, 2002).

Membrane-bound V-PPase activity was first reported in the isolated chromatophores of the photosynthetic, purple, non-sulphur bacterium *Rhodospirillum rubrum* in 1966 (Baltscheffsky *et al.*, 1966), however, it was not until 1989 that the protein itself was purified from plant vacuole membrane fractions of *Vigna radiata* (Maeshima and Yoshida, 1989) and *Beta vulgaris* (Britten *et al.*, 1989; Sarafian and Poole, 1989). V-PPases have now been identified and cloned in a wide range of organisms, including plants, protozoa, bacteria, and archaea. Table 3.1 shows an overview of V-PPases that have been cloned and characterised to date. V-PPase activity has also been reported in a number of other membranes from higher plants, protozoans and bacteria as shown in Table 3.2. Recent BLAST searches have shown the presence of over 100 full-length sequences of putative V-PPases in databases, approximately 50 of those being non-redundant (Au *et al.*, 2006). For a long time, there was no evidence for the existence of V-PPases in human cells or among the rest of the animal kingdom (McIntosh and Vaidya, 2002). However, a recent report showed V-PPase activity in the ovary and egg yolk granules of the insect *Rhodnius prolixus* (Motta *et al.*, 2004), indicating a wider distribution of V-PPases than previously thought.

Recently, it has been shown by biochemical and phylogenetic analysis that V-PPases can be subdivided into two distinct groups based on their dependence on K^+ for enzyme activity (Drozdowicz and Rea, 2001; Belogurov *et al.*, 2002). Type I enzymes are strongly dependent on K^+ (>30 mM) and moderately sensitive to inhibition by Ca^{2+}, whereas Type

II enzymes are fully active in the absence of K^+ and highly sensitive to Ca^{2+} (Drozdowicz *et al.*, 2000; Drozdowicz and Rea, 2001). It has been proposed that K^+-independent V-PPases possess conserved Lys and Thr residues that are absent in K^+-dependent V-PPases. Several reports have described a high homology (85%) among Type I V-PPases in different organisms, whereas only 36% sequence identity exists between Type I and Type II V-PPases (Drozdowicz *et al.*, 2000; Drozdowicz and Rea, 2001; McIntosh *et al.*, 2001). To date, there have been only two reports of the cloning and characterisation of both types of V-PPases in the same organism: *Arabidopsis thaliana* – AVP1 and AVP2 (Sarafian *et al.*, 1992; Drozdowicz *et al.*, 2000), and *Plasmodium falciparum* – PfVP1 and PfVP2 (McIntosh *et al.*, 2001). However, gene sequence database searches have revealed more organisms which contain additional Type I or Type II V-PPase encoding genes (e.g. *Plasmodium yoelii*, McIntosh and Vaidya, 2002).

Table 3.1 Cloned and characterised V-PPases. (adapted from Drozdowicz and Rea, 2001)

Classification			Organism	Reference	Accession number	Name	aa
Eukaryota	**Viridiplantae**						
	Streptophyta	Embryophyta	*Arabidopsis thaliana*	Sarafian *et al.*, 1992	BAA32210	AVP1	770
				Drozdowicz *et al.*, 2000	AAF31163	AVP2	800
			Hordeum vulgare	Tanaka *et al.*, 1993	BAA02717	HVP10	762
				Fukuda *et al.*, 2004	BAB18681	HVP1	771
			Beta vulgaris	Kim *et al.*, 1994b	AAA61609	BVP1	761
					AAA61610	BVP2	765
			Nicotiana tabacum	Lerchl *et al.*, 1995	CAA58701	TVP9	765
					CAA58700	TVP31	766
					CAA54869	TVP5	764
			Oryza sativa	Sakakibara *et al.*, 1996	BAA08232	OVP1	771
					BAA31524	OVP2	767
			Cucurbita moschata	Maruyama *et al.*, 1998	BAA33149		768
			Vigna radiata	Nakanishi and Maeshima, 1998	BAA23649.	VVP	766
			Triticum aestivum	Brini *et al.*, 2005	AAP55210	TVP-1	762
			Suaeda salsa	Guo *et al.*, 2006	-	SsVP	764
		Characeae	*Chara corallina*	Nakanishi *et al.*, 1999	BAA36841	CPP1	793
	Chlorophyta		*Acetabularia acetabulum (mediterranea)*	Ikeda *et al.*, 1999	BAA83103	AcVP	751
	Alveolata						
	Apicomplexa		*Plasmodium falciparum*	McIntosh *et al.*, 2001	AAD17215	PfVP1	717
					AAG21366	PfVP2	1044
			Toxoplasma gondii	Drozdowicz *et al.*, 2003	AAK38076	TgVP1	816

Table 3.1 Cloned and characterised V-PPases. (adapted from Drozdowicz and Rea, 2001) - continued

	Classification	Organism	Reference	Accession number	Name	aa
	Kinetoplastida					
	Trypanosomatidae	*Trypanosoma cruzi*	Hill *et al.*, 2000	AAF80381	TcPPase	*816*
		Trypanosoma brucei	Lemercier *et al.*, 2002	AAK95376	TbVP1	*826*
Archae	**Crenarchaeota**					
	Thermoproteales	*Pyrobaculum aerophilum*	Drozdowicz *et al.*, 1999	AAF01029	PVP	*721*
Bacteria	**α-Proteobacteria**	*Rhodospirillum rubrum*	Baltscheffsky *et al.*, 1998	AAC38615	RVP	*702*
	Thermotogales	*Thermotoga maritima*	Nelson *et al.*, 1999	AAD35267		*726*
	Clostridia	*Carboxydothermus hydrogenoformans*	Belogurov and Lahti, 2002 Wu *et al.*, 2005	ABB14908	ChPPase	*686*
	Actinobacteridae	*Streptomyces coelicolor*	Mimura *et al.*, 2004	BAD36743	ScPP	*794*

Table 3.2 Organisms known to contain V-PPase activity.

Classification		Organism	Reference	Immunolocalisation	Activity assays	Inhibition assays
Eukaryota	**Viridiplantae**					
Streptophyta	Embryophyta	*Zea mays*	Chanson *et al.*, 1985		X	
		Spinacia oleracea	Weiner *et al.*, 1987		X	
Chlorophyta		*Chlamydomonas reinhardtii*	Robinson *et al.*, 1998	X	X	X
Mycetozoa		*Dictyostelium discoideum*	Marchesini *et al.*, 2002	X	X	X
Alveolata						
Apicomplexa		*Toxoplasma gondii*	Rodrigues *et al.*, 2000	X	X	X
Kinetoplastida						
Trypanosomatidae		*Trypanosoma brucei*	Rodrigues *et al.*, 1999a		X	X
		Leishmania donovani	Rodrigues *et al.*, 1999b		X	X
		Leptomonas wallacei	Moraes Moreira *et al.*, 2005	X	X	X
		Herpetomonas species	Soares Medeiros *et al.*, 2005	X	X	X
Metazoa						
Insecta		*Rhodnius prolixus*	Motta *et al.*, 2004	X	X	X
Bacteria	**α-Proteobacteria**	*Agrobacterium tumefaciens*	Seufferheld *et al.*, 2003	X	X	X
	δ-Proteobacteria	*Syntrophus gentianae*	Schocke and Schink, 1998		X	X

3.1.4 Topography of the V-PPase protein and conserved regions in the primary structure

The V-PPase proteins consist of 686 – 1044 aa with molecular weights of 71 – 115 kDa as deduced from the cDNA sequences (Table 3.1). The apparent molecular mass, however, as estimated by SDS-PAGE, ranges from 56 – 79 kDa (Rea and Poole, 1993; McIntosh and Vaidya, 2002). Gel filtration chromatography, radiation inactivation analysis and chemical cross-linking studies demonstrated that V-PPases most likely function as dimers *in vivo* (Sato *et al.*, 1991; Tzeng *et al.*, 1996; Yang *et al.*, 2004), potentially even forming tetramers or hexamers (Mimura *et al.*, 2005). Nevertheless, it has been shown in several studies, utilising heterologous overexpression systems, that a single catalytic subunit of the enzyme is sufficient to convey the enzymatic properties of the V-PPase protein (Kim *et al.*, 1995; Drozdowicz *et al.*, 1999; Hill *et al.*, 2000; Nakanishi *et al.*, 2001; Belogurov and Lahti, 2002; Ikeda *et al.*, 2002b; Hirono *et al.*, 2005; Lin *et al.*, 2005). Analysis of hydropathy plots in conjunction with topological prediction methods have shown that the single V-PPase polypeptide consists of 14-17 α-helical transmembrane domains (TMDs) (McIntosh and Vaidya, 2002; Mimura *et al.*, 2004). To date, there have been no reports on the tertiary structure of the V-PPase protein due to the difficulty of crystallising this highly hydrophobic membrane protein. However, Mimura *et al.* (2005) confirmed the predicted topological arrangement of TMDs in *S. coelicolor* by cysteine-scanning mutagenesis. Figure 3.6 shows a diagram of the proposed membrane topology of ScPP, displaying 17 TMDs (Mimura *et al.*, 2004). As *S. coelicolor* exhibits a long C-terminal tail that is not common amongst most V-PPases, it was proposed that all other V-PPases lacking this additional sequence have 16 TMDs, thus placing both the amino (N)-terminus and carboxy (C)-terminus in the vacuolar lumen (Mimura *et al.*, 2004).

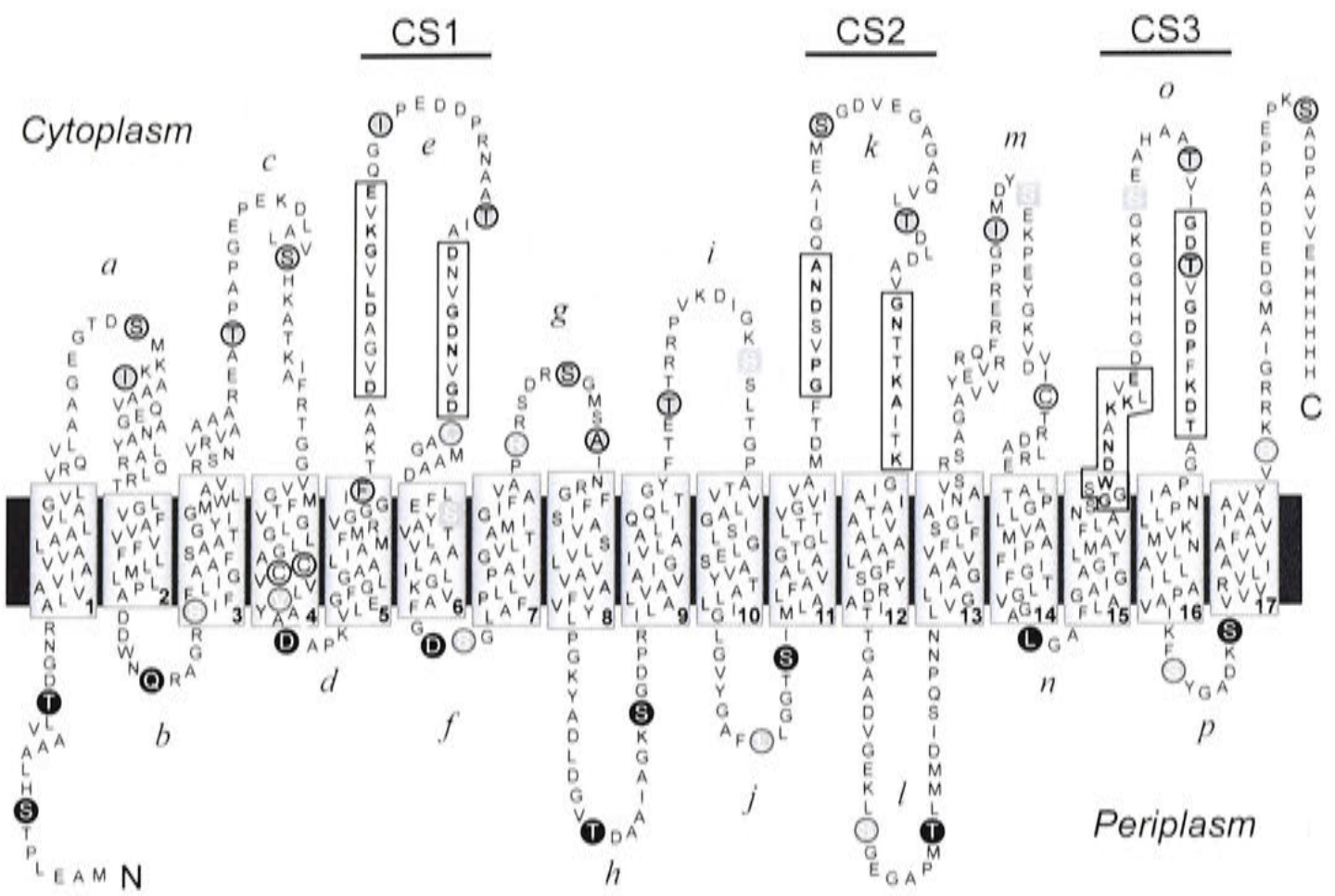


Figure 3.6 Model of the membrane topology of ScPP. The TMDs (1–17) of ScPP were predicted with the TMHMM program (http://www.cbs.dtu.dk/services/TMHMM). Conserved motifs such as DVGADLVGKVE are boxed. Residues common to various V-PPases are in boldface, and the conserved segments (CS1, CS2, and CS3 - see text below for further detail) are depicted. The six histidine residues at the C-terminus are the His6 tag used for purification of the overexpressed protein. (Source: Mimura *et al.*, 2004)

Multiple sequence comparison of a great number of membrane-bound V-PPase protein sequences revealed a number of highly conserved regions, in particular the three segments CS1, CS2 and CS3 located in the cytosolic loops *e*, *k* and *o*, respectively, as shown in Figure 3.6 (Mimura *et al.*, 2004). Several structure-function studies of the V-PPase protein have been carried out in the past 10 years to identify the functional domains involved in enzyme activity. The catalytic domain for substrate hydrolysis is thought to be composed of at least three conserved motifs (Baltscheffsky *et al.*, 1999; Nakanishi *et al.*, 2001; Mimura *et al.*, 2004). The putative substrate-binding sequence Dx_7KxE is located in segment CS1. Its involvement in PP_i hydrolysis and H^+-translocation activity has been shown by the loss of specific enzyme activity in immunochemical studies using antibodies specific to the catalytic region (Takasu *et al.*, 1997) and by site-directed mutagenesis (Nakanishi *et al.*, 2001). The three charged residues of the acidic motif Dx_3Dx_3D, located in segments CS1 and CS3, have also been shown to be involved in enzyme activity (Nakanishi *et al.*, 2001). In addition, it was demonstrated that a specific antibody against an 11 aa sequence of the C-terminus of the *V. radiata* V-PPase was able to suppress hydrolytic and H^+-translocation activities,

suggesting a location in proximity and/or interaction with the catalytic site (Takasu *et al.*, 1997). These findings were supported by a recent deletion mutation study, where a series of C-terminal truncated *V. radiata* V-PPase proteins was analysed for enzyme activity. Lin *et al.* (2005) demonstrated that the deletion of the last five amino acids in the C-terminus led to a drastic loss in enzyme activity and proton translocation, hence it was concluded that the C-terminal region plays an essential role in the function of the enzyme. A recent study furthermore suggested the involvement of TMD5, another highly conserved region located at the N-terminal side of cytosolic loop *e*, in proton translocation of V-PPases. TMD5 shows a lower degree of hydrophobicity compared to other domains and site-directed mutagenesis revealed a GYG motif which presumably plays a role in sustaining the enzymatic reaction of V-PPases (Van *et al.*, 2005). It has been concluded that V-PPase activity involves three conserved motifs that form the catalytic site and requires conformational changes that affect at least seven cytoplasmic loops: four loops containing conserved motifs (*e*, *k*, *m*, and *o*) and three other loops (*c*, *g*, and *i*) (Mimura *et al.*, 2005).

A number of motifs used as criteria for distinguishing Type I and Type II V-PPases based on their sequence, have recently been described. It was shown that an alanine to lysine substitution in position 460 of the *C. hydrogenoformans* V-PPase is sufficient to confer K^+ independence to both PP_i hydrolysis and PP_i-energized H^+ translocation (Belogurov and Lahti, 2002). Therefore, the motif GNxx[A/K]Ax[TA/G] located in the cytoplasmic loop *k* between TMD11 and TMD12 (CS2) can be used to classify uncharacterised V-PPases based on available sequence information (see Figure 3.7). Another TMD/loop motif, which is [DE]YYTS in all Type I V-PPases, is replaced by [KQR]YYT[DE] in Type II V-PPases. Interestingly, it was found that the glutamic acidic residue (E) present in all Type I V-PPases is essential for H^+-translocation. Site-directed mutagenesis experiments on AVP1 which generated mutants bearing either a glutamine or lysine instead of the glutamic acid in position 427, rendered the enzyme incapable of H^+-translocation (Zhen *et al.*, 1997). However, it was shown that heterologously expressed AVP2, a Type II V-PPase carrying a residue other than glutamic acid in position 427, is indeed competent in both PP_i hydrolysis and PP_i-energized H^+-translocation. It was therefore proposed that Type II V-PPases are competent in H^+-translocation because a second substitution (Ser to Asp) three places displaced from this position introduces an acidic residue that can serve the same function as E427 (Drozdowicz *et al.*, 2000).

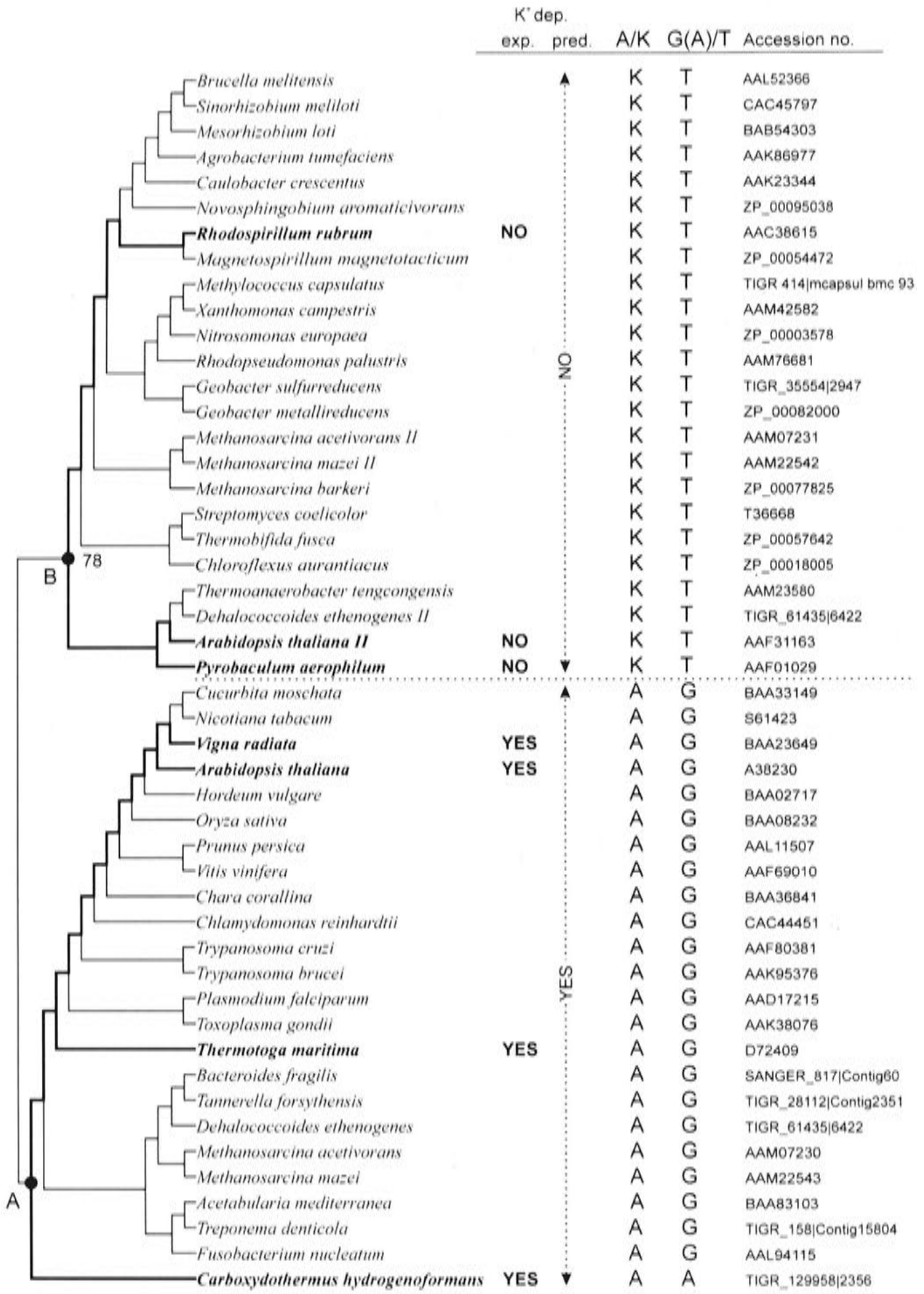


Figure 3.7 K^+ requirement of functionally characterised V-PPases and other available V-PPase sequences on the basis of phylogenetic analyses. A multiple sequence alignment was generated by ClustalX, and subjected to phylogenetic analysis using both maximum likelihood and distance (neighbor-joining) methods (PHYLIP). Functionally characterized V-PPases are indicated in boldface, and their K^+ requirement is traced back along the branches (boldface lines) to nodes A and B that are accentuated by black circles. For the node B, bootstrap value is indicated as percentage of 1000 replications. The columns to the right of the tree display K^+ requirement of functionally characterized V-PPases (K^+ dep., exp. column), inferred K^+ requirement of all H+-PPase sequences (K^+ dep., pred. column), residues in Ala/Lys (A/K column) and Gly(Ala)/Thr (G(A)/T column) positions in the multiple sequence alignment and protein sequence accession numbers in the GenBank or DNA contig numbers for preliminary sequences obtained from The Institute for Genomic Research (www.tigr.org) and The Sanger Center (www.sanger.ac.uk). (Source: Belogurov and Lahti, 2002)

3.1.4.1 Functional residues

Substances that are commonly used in enzyme studies to elucidate potential binding sites and to identify essential amino acid residues involved in enzymatic and H^+-translocating reactions, often in combination with site-directed mutagenesis, are summarised in Table 3.3 (reviewed in Schultz and Baltscheffsky, 2004).

A number of studies over the past years have revealed some important residues potentially involved in V-PPase enzyme activity. Radioactive *N,N'*-dicyclohexylcarbodiimide (DCCD) labelling/peptide mapping demonstrated that residue D283 (*V. radiata*) is the sole target for DCCD inhibition, indicating its involvement in the enzyme reaction (Yang *et al.*, 1999). Furthermore, the two acidic residues E305 and D504 from *A. thaliana* AVP1 have been shown to directly participate in DCCD-binding, and it was proposed that both residues are critical for catalysis of the V-PPase enzyme, as amino acid substitutions resulted in a loss of both PP_i hydrolysis and H^+-translocation (Zhen *et al.*, 1997). Moreover, a *N*-ethylmaleimide-binding (NEM) cysteine residue, C634 in AVP1 from *A. thaliana*, has been shown to be crucial for inhibition of the V-PPase enzyme by maleimides but not for catalysis (Kim *et al.*, 1995). A more recent amino acid substitution study in *R. rubrum* identified two residues potentially involved in substrate binding (E197 and E202) and two residues which are deemed to be important for the correct folding of the polypeptide (E550 and E649) (Malinen *et al.*, 2004). The histidine-modifying reagent DEPC was used to demonstrate the importance and possible involvement of histidine residue H716 (*V. radiata*) in the activity of V-PPases (Hsiao *et al.*, 2004). Furthermore, inhibition of the V-PPase with fluorescein 5'-isothiocyanate (FITC) indicated that the residue targeted by the reagent might be located in or near the catalytic domain of the enzyme (Yang *et al.*, 2000). A recent study identified the residue as K469 (*R. rubrum*) by a combination of site-directed mutagenesis and FITC inhibition, and a substrate-binding involvement of the residue has been proposed (Schultz and Baltscheffsky, 2004).

Table 3.3 Components commonly used for protein studies

Substance	Abbrev.	Site of activity	Reference
N,N'-dicyclohexylcarbodiimide	DCCD	Acidic residues	Zhen *et al.*, 1997
Fluorescein 5'-isothiocyanate	FITC	Amines/SH	Yang *et al.*, 2000
Diethyl pyrocarbonate	DEPC	His (H)	Hsiao *et al.*, 2002
N-ethylmaleimide	NEM	SH groups (Cys – C)	Kim *et al.*, 1995

3.1.4.2 V-PPase and its substrate – the role of Mg^{2+}

V-PPase specifically utilises PP_i and requires free Mg^{2+} as a cofactor, as the binding of Mg^{2+} stabilises and activates the enzyme (Gordon-Weeks *et al.*, 1996; Maeshima, 2000). It is thought that a single V-PPase subunit, like the sPPases, contains at least two distinct binding sites for Mg^{2+} (Cooperman, 1982; Baykov *et al.*, 1993; Maeshima, 2000; Nakanishi *et al.*, 2001). These proposed sites show different affinities towards the Mg^{2+} ion: a low affinity site with a K_m value of 0.25-0.46 mM, and high affinity site with a K_m value of 23-31 μM (Baykov *et al.*, 1993; Gordon-Weeks *et al.*, 1996).

The enzyme uses a Mg–PP_i complex as a substrate, however, there has been much dispute in the past over whether $MgPP_i$ or Mg_2PP_i is the active substrate (Rea and Poole, 1993; Baykov *et al.*, 1996; Gordon-Weeks *et al.*, 1996). Protection studies and more recent amino acid substitution studies indicate that Mg_2PP_i is in fact the substrate used by the V-PPase protein. It was shown that the protein was protected from chemical inactivation or protease degradation in the presence of Mg^{2+} and Mg_2PP_i, whereas high concentrations of MgPPi or PP_i alone did not provide any protection (Maeshima, 1991; Gordon-Weeks *et al.*, 1996; Nakanishi *et al.*, 2001). Furthermore, binding of Mg^{2+} and Mg_2PP_i also protects the enzyme from heat inactivation (Maeshima, 1991; Yang *et al.*, 2004). In a recent protection and substitution study it was shown that substitution of residue D217 of the *R. rubrum* V-PPase protein leads to a change in the substrate preference of the enzyme from Mg_2PP_i to $MgPP_i$ (Schultz and Baltscheffsky, 2004), thus confirming Mg_2PP_i as the principal substrate for V-PPases.

3.1.5 Enzyme regulation and inhibition of V-PPase activity

V-PPase activity is regulated by a number of inhibitors and activators, such as mono- and divalent cations and free PP_i. As described above, Mg^{2+} is an essential co-factor for V-PPase enzyme activity. Furthermore, Type I V-PPases also require the presence of K^+. Other monovalent cations such as Rb^+ and NH_4^+ are able to substitute for K^+ as an activator with the same degree of effectiveness on enzyme activity (Walker and Leigh, 1981). Excess concentrations of free PP_i, Na^+, Ca^{2+} and F^- on the other hand inhibit V-PPase enzyme activity. Free Ca^{2+} has been shown to specifically inhibit V-PPase activity by direct interaction with the enzyme (Rea *et al.*, 1992; Rea and Poole, 1993) and by preventing the formation of the substrate complex Mg_2PP_i by chelating PP_i (Maeshima, 1991). Inhibitory effects of Na^+ on V-PPase enzyme activity was demonstrated on

vacuolar membrane fractions from *B. vulgaris*, indicating that the inhibition effectiveness of Na^+ salts is dependent upon the associated anion (Walker and Leigh, 1981; Rea and Poole, 1985). The effect of substrate inhibition (free PP_i), common to most types of PPases, was described by Leigh *et al.* (1992). Therefore, in addition to a substrate-binding site and a H^+ conducting channel, V-PPases are likely to have additional binding sites for Mg^{2+}, Ca^{2+} and in some instances monovalent cations such as K^+ (McIntosh and Vaidya, 2002). However, to date no essential regions for those sites have been determined.

Until the mid-1990s there was no known type-specific inhibitor of V-PPases available. All other classes of H^+-translocating enzymes were easily distinguished from the others on the basis of susceptibility to inhibition by type-specific inhibitors (e.g. F-ATPase – azide, P-ATPase – orthovanadate, V-ATPase – bafilomycin) (Zhen *et al.*, 1994a). A range of artificial substances was subsequently tested in order to identify a potent and highly specific inhibitor of the V-PPases. It was found that the family of 1,1-bisphosphonates, analogues of the substrate PP_i, competitively inhibit soluble and membrane-bound PPases with varying degrees of effectiveness and specificity (Zhen *et al.*, 1994a; Gordon-Weeks *et al.*, 1999). In this family, the central oxygen atom of the PP_i is replaced with a carbon atom which can accommodate two additional covalent bonds, hence giving rise to a variety of compounds (McIntosh and Vaidya, 2002). The compound aminomethylenebisphosphonate (AMBP, Figure 3.8) was shown to be one of the most potent and specific inhibitors of V-PPases *in vitro*. It does not inhibit ATPases or other non-V-PPase related PP_i hydrolysis (Zhen *et al.*, 1994a; Gordon-Weeks *et al.*, 1999). Several *in vivo* inhibition studies with *Plasmodium*, *Toxoplasma* and *Trypanosoma* species demonstrated sensitivity to AMBP and subsequent growth inhibition of cultured organisms (reviewed in McIntosh and Vaidya, 2002). However, it was noted that the efficiency of AMBP was significantly lower *in vivo* than *in vitro* (Drozdowicz *et al.*, 2003), potentially due to bulky anionic character of bisphosphonates and the inefficient uptake by the organism (McIntosh *et al.*, 2001).

Pyrophosphate (PP_i)

$H_2O_3P - O - PO_3H_2$

Aminomethylenebisphosphonate (AMBP)

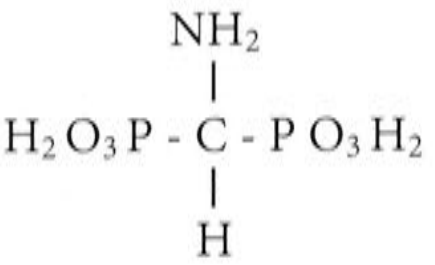


Figure 3.8 Structure of pyrophosphate and AMBP. (adapted from McIntosh and Vaidya, 2002)

3.1.6 Potential role of V-PPase in osmoregulation

Contractile vacuoles (CV) of numerous organisms have been shown to be dependent on V-ATPase function (reviewed in Allen, 2000; Allen and Naitoh, 2002; Becker and Hickisch, 2005). The exact mechanism by which the CV expels water from the cytoplasm remains unclear to date, however, it has been suggested that the V-ATPase drives the establishment of a proton gradient in the membrane by H^+-translocation which in turn facilitates antiport of cytosolic osmolytes with following passive or osmotic influx of water (Nolta and Steck, 1994; Marchesini *et al.*, 2002). Recently, a number of studies located the V-ATPase and V-PPase to the same membrane system. Co-localisation of these enzymes was observed in the acidocalcisomes and the CV of *C. reinhardtii* (Ruiz *et al.*, 2001a), the acidocalcisomes of apicomplexan and trypanosomatid parasites (reviewed in Docampo and Moreno, 2001; Rohloff *et al.*, 2004), the mass-dense granules and the CV of *D. discoideum* (Marchesini *et al.*, 2002), the digestive vacuole of *P. falciparum* (Saliba *et al.*, 2003), plant vacuoles (reviewed in Maeshima, 2001), and the yolk granule membranes of *R. prolixus* (Motta *et al.*, 2004). A link between the mass-dense granules/acidocalcisomes of *D. discoideum* and *C. reinhardtii*, and the acidocalcisomes of protozoan parasites and their involvement in osmoregulation has recently been proposed (Docampo and Moreno, 2001; LeFurgey *et al.*, 2001; Ruiz *et al.*, 2001b). It is thought that V-PPases might contribute to an additional driving force for water uptake (Marchesini *et al.*, 2002).

As described in Chapter 1 (1.2.2.2), CV complexes are thought to be composed of two distinctly different membrane populations, the spongiome and the bladder. The spongiome is often divided into numerous vesicles and tubules, and is rich in V-ATPases, whereas the membrane of the bladder lacks V-ATPases but contains a V-PPase. The spongiome is thought to be capable of fusion with only the bladder membrane system, while the bladder can also fuse with the plasma membrane (reviewed in Allen and

Naitoh, 2002). In *P. nicotianae* zoospores, Mitchell and Hardham (1999) showed that the bladder shares a P-type ATPase with the plasma membrane, while the spongiome lacks the P-type ATPase but contains a V-ATPase instead. This conforms with the proposed two-compartment model. However, recent studies using subcellular fractionation revealed that isolated CV bladders of *T. cruzi* possess V-ATPase and V-PPase activity (Rohloff *et al.*, 2004).

3.1.7 Aims of this study

This chapter aimed to increase our understanding of V-PPases in *Phytophthora* by the isolation and characterisation of the V-PPase gene from a *P. nicotianae* zoospore cDNA library (*PnVPP*). In order to determine PnVPP enzyme activity, a phosphatase activity assay was carried out using microsomal fractions of *P. nicotianae* zoospores. The V-PPase type-specific inhibitor AMBP was furthermore used to examine its effect on PnVPP activity in *P. nicotianae in vitro* and *in vivo*. It was also aimed to localise the PnVPP protein in P. *nicotianae* zoospores by the production of monoclonal and polyclonal antibodies.

3.2 Results

3.2.1 Analysis of cDNA clone DS311

The cDNA clone DS311 was isolated from a *P. nicotianae* λ-ZAP zoospore library by Dr Dubravka Škalamera during the differential screening of randomly selected clones. Plasmid DNA was isolated from the clone DS311 and was sequenced from the 5'-end using the T3 pBS primer as well as primer DS311_For1, and from the 3'-end using pBS primer T7. DS311 encodes 396 aa (reading frame 3). Protein-Protein BLAST analysis (BLASTP) of the translated protein sequence showed similarity with a H^+-pyrophosphatase in the CDD (CDD Access Number 26008 pfam03030; Figure 3.9 and Figure 3.10). Further GenBank analysis (translated query vs. protein database - BLASTX) of the 1366 bp insert showed homology to a vacuolar-type H^+-translocating inorganic pyrophosphatase in a wide range of organisms. The most similar was the *A. thaliana* 'Vacuolar-Pyrophosphatase Like Protein 1' (AVPL1 - GenBank accession number NP_173122) with an *E* value of 2×10^{-133} indicating a high level of homology (Figure 3.11). Significant hits were also observed against proteins in *O. sativa* (GenBank accession number BAD27918), several *Plasmodium* species (e.g. *P. falciparum*, NP_701702) and a wide range of bacteria. The length of these homologous proteins ranged from about 700 – 1000 aa, indicating that clone DS311 does not contain the full length cDNA for the protein in *P. nicotianae*. The protein encoded by cDNA clone DS311 was named PnVPP for *Phytophthora nicotianae* Vacuolar PyroPhosphatase.

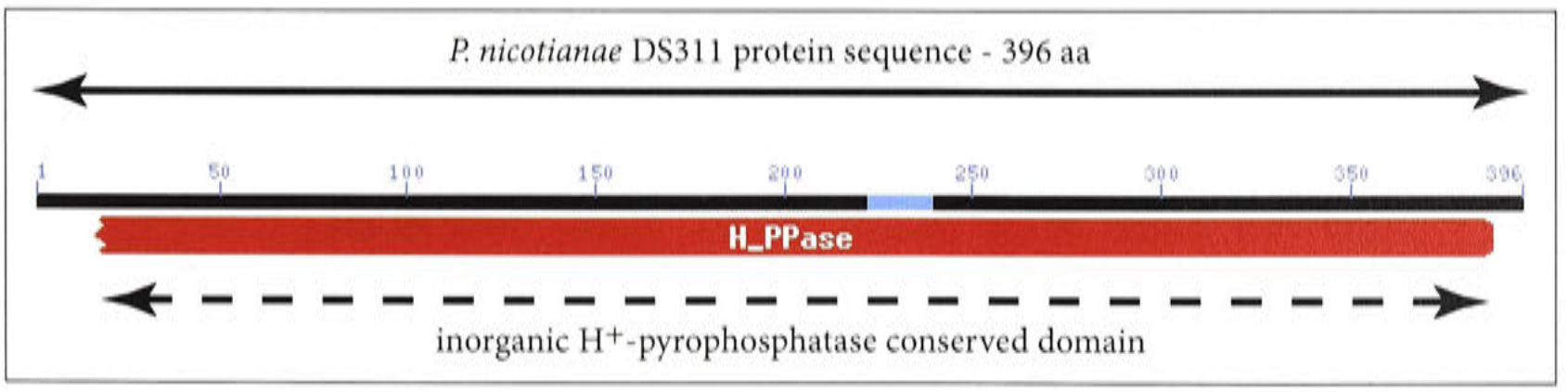


Figure 3.9 Conserved domain of the inorganic H^+-pyrophosphatase family obtained by BLASTP search using the inferred amino acid sequence of cDNA clone DS311.

```
                CD-Length = 682 residues, only  50.3% aligned
                Score =  374 bits (961), Expect = 1e-104

Query:   16   FFLCGVVGMLTAYVFVKSTQYYTDYAHPPVRSIAKASTTGHGTNIITGVAVGMKSTVVPT  75
Sbjct:   339  ALVAVLIGLWAGLLIGFITEYYTSYNYRPVRDIAKSSKTGAATNIIAGLSLGLKSTAIPI  398

Query:   76   LMVSFAVISAYHLGASSGIGGVGNRHAGLFGTAVATMGMLSSAVFVLAMNNYGPIADNAG  135
Sbjct:   399  LTIAAAIIIAYYL------------AGLYGIAIAAVGMLSTAGIILAIDAYGPISDNAG  445

Query:   136  GIAEMSRQPDYVRDATDKLDAAGNVTKAITKGYSIGSAALACFVLFGAFMDEFSEFAGRE  195
Sbjct:   446  GIAEMAGLPEEVREITDALDAVGNTTAAIGKGFAIGSAALAALALFGAYMDAADI-----  500

Query:   196  FKTVDIATVEVLVGGLLGTMMVFFFTGLAVAAVGETAGEVVNEVRHQFEIYPGIMEYKAK  255
Sbjct:   501  -GEVNILNPNVIVGLLLGAMLPYLFSALTMKAVGRAAMEVVEEVRRQFREIPGIMEGKEK  559

Query:   256  PDYRTCVALVTEAALKQMRFPGLLAVLMPVSVGIIFRVIGEYQGKPLLGAEALAGYLMFG  315
Sbjct:   560  PDYGRCVDISTDAALREMIIPGLLAVLTPLVIGF------------LLGAEALGGLLAGA  607

Query:   316  TVTGIMMALFLDNVGGAWDNAKKYVELG-NFGGKGSEAHKAAVTGDTVGDPFKDTAGPAL  374
Sbjct:   608  IVSGVLMAISLANSGGAWDNAKKYIEAGGHEGGKGSEAHKAAVIGDTVGDPLKDTSGPSL  667

Query:   375  HVVIKLLSTTVLVF  388
Sbjct:   668  NILIKLMAVVSLVF  681
```

Figure 3.10 Amino acid sequence alignment of cDNA clone DS311 with the H⁺-pyrophosphatase conserved domain (CD) from the NCBI database. The DS311 amino acid residues from position 16-388 aligned with the H⁺-pyrophosphatase conserved domain amino acid residues from 339-681 and produced an overall alignment of 50% between the two sequences. The residues in red are identical in the two sequences whereas those in blue represent substitutions with conserved residues.

```
gi|15219362|ref|NP_173122.1|
 AVPL1 (VACUOLAR-PYROPHOSPHATASE LIKE PROTEIN 1) [Arabidopsis thaliana]

Length=802

Score =  479 bits (1232),  Expect = 2e-133
Identities = 247/389 (63%), Positives = 311/389 (79%), Gaps = 1/389 (0%)
Frame = +3

Query  18    TPEYPSAWMNFFLCGVVGMLTAYVFVKSTQYYTDYAHPPVRSIAKASTTGHGTNIITGVA  197
             T + PSAW+NFF+CG+VG++TAYVFV  ++YYTDY + PVR++A AS+TGHGTNII GV+
Sbjct  414   TEQAPSAWLNFFMCGLVGIITAYVFVWISRYYTDYKYEPVRTLALASSTGHGTNIIAGVS  473

Query  198   VGMKSTVVPTLMVSFAVISAYHLGASSG-IGGVGNRHAGLFGTAVATMGMLSSAVFVLAM  374
             +G++ST +P L++S A+ISA+ LG +SG I   GN   GLFGTAVATMGMLS+A +VL M
Sbjct  474   LGLESTALPVLVISVAIISAFWLGNTSGLIDEKGNPTGGLFGTAVATMGMLSTAAYVLTM  533

Query  375   NNYGPIADNAGGIAEMSRQPDYVRDATDKLDAAGNVTKAITKGYSIGSAALACFVLFGAF  554
             + +GPIADNAGGI EMS+QP+ VR+ TD LDA GN TKA TKG++IGSAALA F+LF A+
Sbjct  534   DMFGPIADNAGGIVEMSQQPESVREITDVLDAVGNTTKATTKGFAIGSAALASFLLFSAY  593

Query  555   MDEFSEFAGREFKTVDIATVEVLVGGLLGTMMVFFFTglavaavgetagevvnevRHQFE  734
             MDE S FA   FK VDIA  EV +GGLLG M++F F+  A AAVG TA EVVNEVR QF
Sbjct  594   MDEVSAFANVSFKEVDIAIPEVFIGGLLGAMLIFLFSAWACAAVGRTAQEVVNEVRRQFI  653

Query  735   IYPGIMEYKAKPDYRTCVALVTEAALKQMRFPGLLAVLMPVSVGIIFRVIGEYQGKPLLG  914
               PGIM+YK KPDY  CVA+V  +AL++M  PG LA++ P++VG +FR++G Y G+PLLG
Sbjct  654   ERPGIMDYKEKPDYGRCVAIVASSALREMIKPGALAIISPIAVGFVFRILGYYTGQPLLG  713

Query  915   AEALAGYLMFGTVTGIMMALFLDNVGGAWDNAKKYVELGNFGGKGSEAHKAAVTGDTVGD  1094
             A+ +A  LMF TV GI+MALFL+  GGAWDNAKKY+E G  GGKGS++HKAAVTGDTVGD
Sbjct  714   AKVVAAMLMFATVCGILMALFLNTAGGAWDNAKKYIETGALGGKGSDSHKAAVTGDTVGD  773

Query  1095  PFKDTAGPALHVVIKLLSTTVLVFGPLFV  1181
             PFKDTAGP++HV+IK+L+T  LV  P+F+
Sbjct  774   PFKDTAGPSIHVLIKMLATITLVMAPIFL  802
```

Figure 3.11 Best GenBank homology BLASTX alignment result of the *P. nicotianae* cDNA clone DS311. The 1366 bp insert of the cDNA clone shows high homology to an *A. thaliana* inorganic H⁺-pyrophosphatase.

3.2.2 RNA Blot Expression Analysis of cDNA clone DS311

To characterise the expression profile of *PnVPP* in *P. nicotianae*, RNA blot analysis was carried out. The RNA blot was prepared from total RNA of four asexual life cycle stages of *P. nicotianae*: vegetative hyphae, 4 h sporulating hyphae, zoospores and 2 h germinated cysts (isolated by Val Maclean). The ^{32}P-labelled DS311 probe revealed a transcript of ~ 2.4 kb, confirming that the *PnVPP* clone does not contain the full length cDNA (Figure 3.12). The *PnVPP* transcript is upregulated in zoospores and germinated cysts but not detected in sporulating hyphae. Equal loading of RNA was demonstrated by methylene blue staining of rRNAs and hybridisation with the constitutively expressed cDNA clone WS41 (data not shown, Shan *et al.*, 2004).

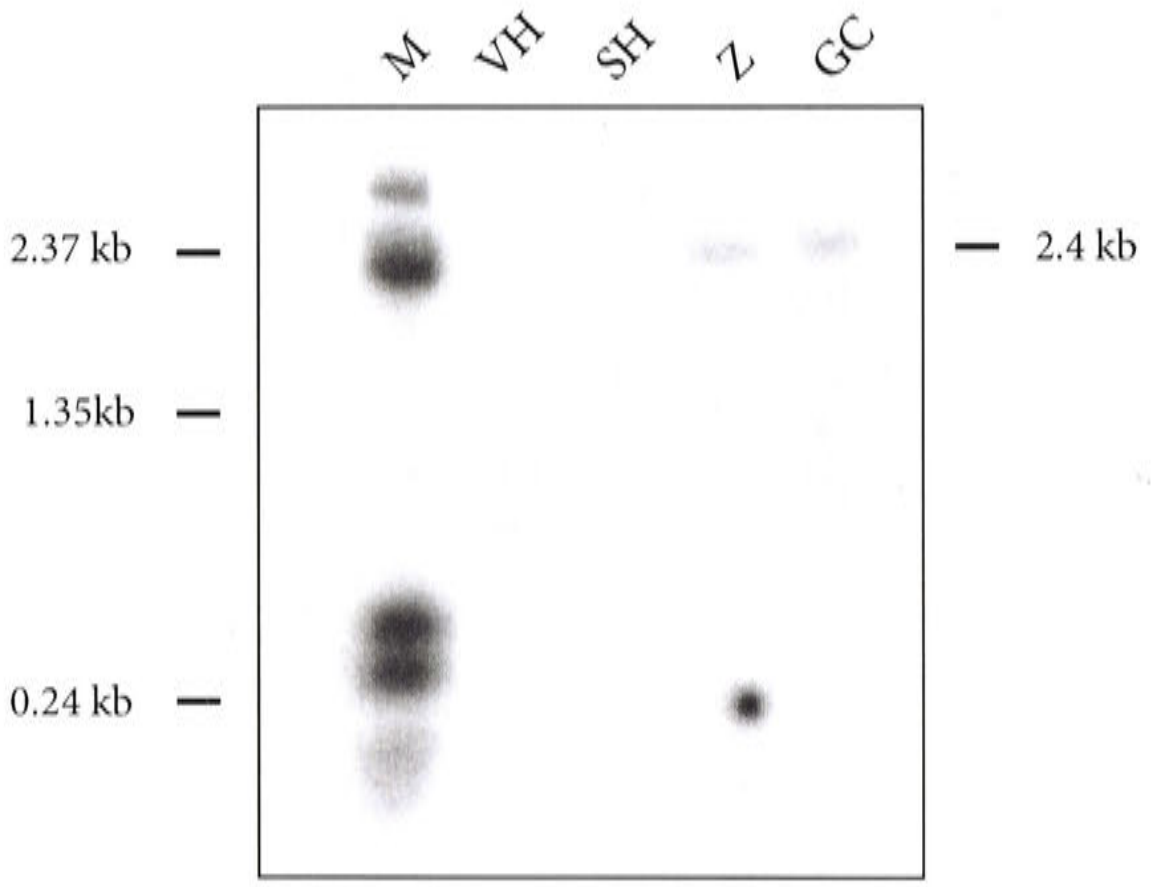


Figure 3.12 Northern blot analyses with RNA from four life cycle stages of *P. nicotianae* probed with ^{32}P-labelled cDNA clone DS311. The hybridisation result indicates expression of the ~2.4 kb *PnVPP* transcript in vegetative hyphae (VH), zoospores (Z) and germinated cysts (GC) but not in sporulating hyphae (SH). Expression is up-regulated in zoospores and germinated cysts. M- size marker

3.2.3 In search of the 5' end: Screening of a zoospore λ-library and 5' RACE

As the alignment of the available cDNA sequence and the genomic clone indicated two potential ATG start sites (see Section 3.2.4), the full length cDNA sequence had to be obtained in order to identify the correct translation start site. The λ-ZAP zoospore library was re-screened with a probe situated at the first in-frame ATG start codon. A 671 bp product was PCR amplified from genomic *P. nicotianae* DNA with primers probe_5'end for and probe_5'end rev. The purified and ^{32}P-labelled product (5'end probe) was used to screen the library as described in Section 2.2.7. In the primary screening of the library five putative positive clones were identified. The second round of screening revealed that only one of the five putative clones was positive (Figure 3.13). Six phage colonies of the positive clone (named A-G) were picked and the pBS phagemids, containing the cDNA inserts, were excised from the vector. Several colonies obtained from each of the six individual excisions were subjected to colony PCR to confirm the correct insert (data not shown). Due to inconsistent results, Southern blot analysis was carried out after double digestion of the plasmid DNA with *Eco*RI and *Xho*I. Clones showed varying band patterns after digest and only four out of nine clones reacted with the probe (Figure 3.14). Sequencing of three clones (A-5, E-4 and E-20) revealed that clones A-5 and E-20 contain the correct insert. Full sequencing of the inserts and alignment with the genomic DNA and the original cDNA clone DS311 showed that an additional 750 bp of cDNA information was obtained.

However, as approximately 500 – 700 bp were still unaccounted for (Figure 3.15), a 5' RACE was performed in order to obtain the missing cDNA information. RNA was isolated from zoospores and subjected to 5' RACE as described in Section 2.4.3 using a set of three nested primers (Figure 3.16). The experiment was repeated three times and resulted in a 500 bp band in all individual runs. Negative controls lacking RT in the cDNA synthesis step did not yield any bands. The 500 bp band obtained (Figure 3.17) was subcloned into a pGEM T-easy vector and fully sequenced.

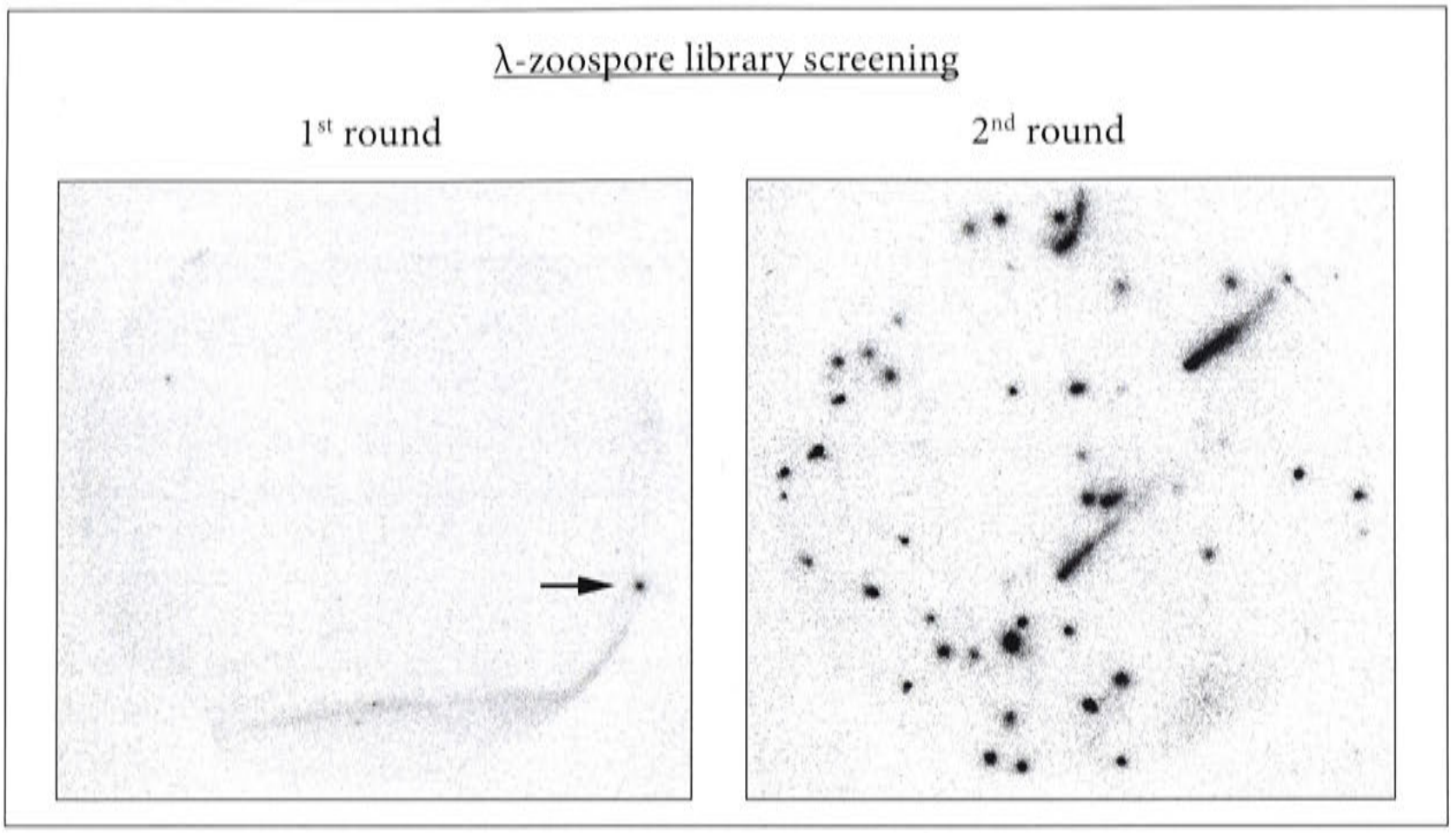


Figure 3.13 Hybridisation of membranes containing putative *PnVPP* cDNA clones from a *P. nicotianae* zoospore library with ^{32}P-labelled 5'end-probe. Hybridisation results of one positive clone (arrow) after two rounds of screening are shown.

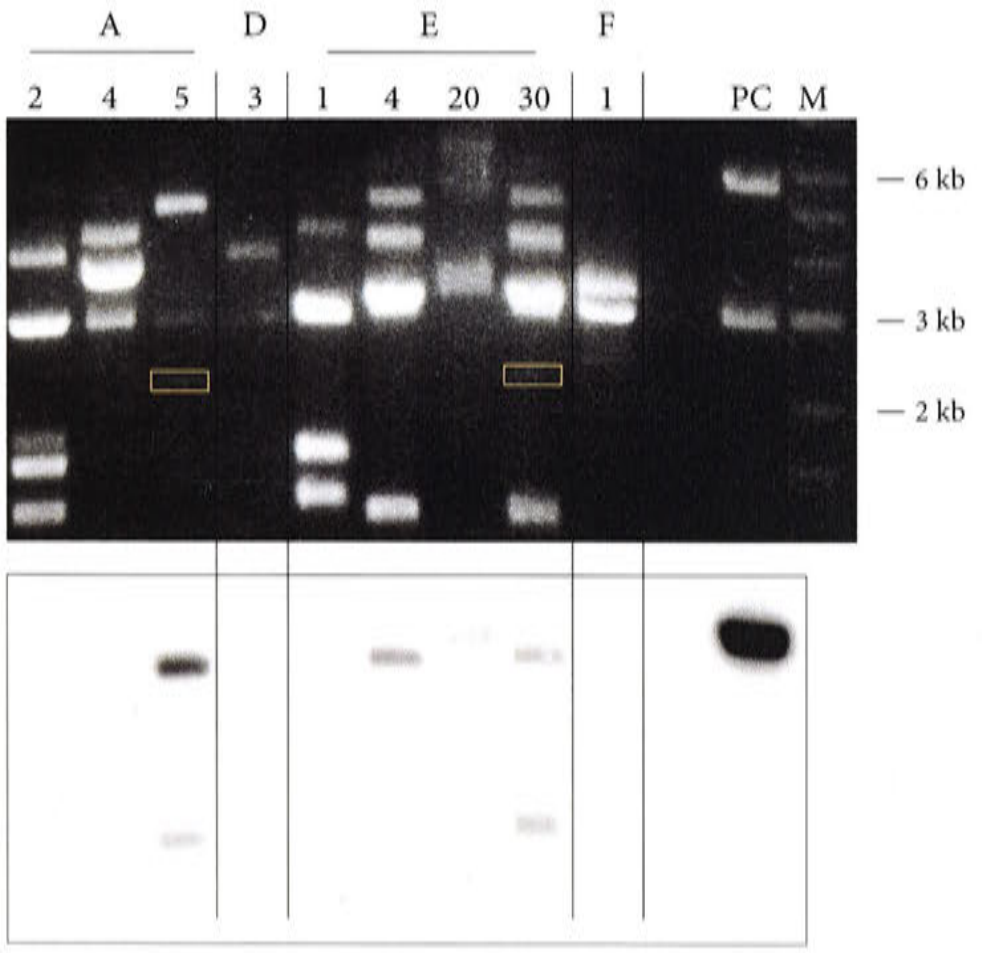


Figure 3.14 *Eco*RI/*Xho*I restriction digest of plasmid DNA from putative *PnVPP* cDNA clones and hybridisation with ^{32}P-labelled 5'end probe. Letters A-F represent the six clones that were excised after the second round of screening. Four out of the nine clones that initially tested positive in colony PCR reacted with the probe. Plasmid DNA was only partially digested, therefore two bands are labelled. No digestion of plasmid DNA was observed in clone E-20. The lower band (~2.2 kb, orange box in the EtBr stained gel picture) represents the released insert, whereas the upper band shows the undigested DNA (clones A-5, E-4 and E-30). The genomic clone 21K17 (3.2.4) used as a positive control (PC) shows a strong reaction with the probe.

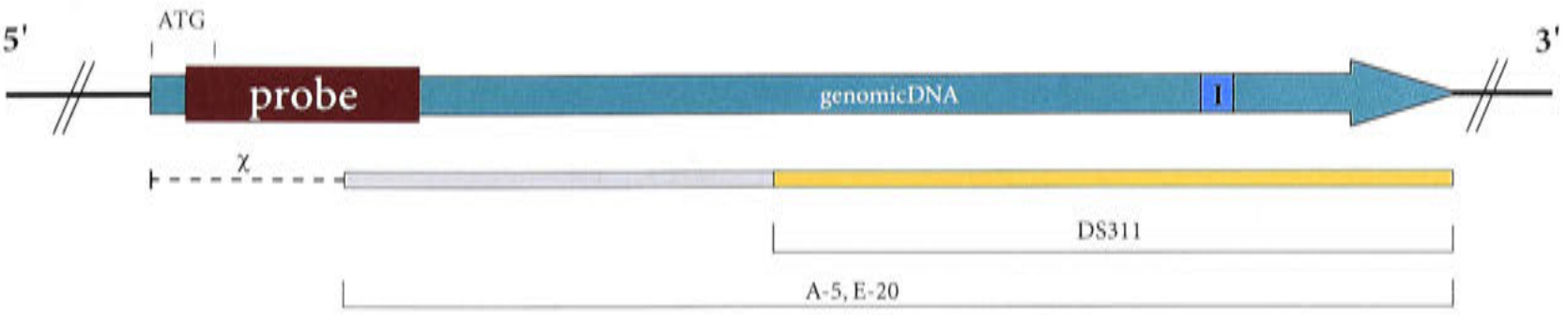


Figure 3.15 A schematic diagram of the *PnVPP* gene. The green area depicts the genomic DNA sequence of the gene, 'I' (blue box) indicates a single intron. The hybridisation experiments used a probe (red box) that was situated at the potential ATG start site as inferred from the genomic sequence (3.2.4). Available cDNA sequence information is shown below the genomic DNA. The orange box denotes the original cDNA clone DS311 sequence. The combined orange box and grey boxes represent cDNA clones A-5 and E-20, with the grey box showing cDNA sequence information in addition to that from DS311. The dashed line indicates the region of missing cDNA information (χ).

TTTGATAGGATGCACCAATCAGGAACGTTTCTAGGACAGCTGGTTGAAGTTACCTCGTTATAGTGACAGT

TTAGTCATTTCAAGTCAGTCACTATTGCAAGGTAGAAAACAGCAGGTCAGATTCTGATCCGCACGCGTGT

CAGTCAAAGAGCAGCAATCGAGAGGTCGGACCGTTCACGTCGACGCCAGCTACGTGATCCCCACTGCTTT

CCGGTGCTGCACTCCGCTCTCTCTGTTCCCAAGGCATTCGTTCCGTGCTCTCCGCTTCTCTGAACTCCGC

CTACAGCCAGCGTCATGCACGAGAAGACACAGGTTGACATCGAGTTCGGCATGTCGGACACTGCACCGCT

GCACAGTGAGCCAGACCTCGAAGAAATGGGGTCGGGTGCGTCTGCGGCCTTGGCTTACGCCCGTACACTG

GATCTACAGCCTCCGCTCATCGGTATCTCGGTGATAGGAGCGCTGCTGCTGCTGTTCGGTGGCAGCTTCC

GCCCCATCTTTCTGGTCTTCACTATCTGCGGCTACGGCTCCGTGTTCGCGCTCTACCTGTCACACTGGGT

GCTATCCAAGGACGCAGGCACGACCGAGATGAAGGAAGTAGCCACGCCCATCCGTCAGGGAGCCGAGGGC

GSP3

TTTCTCAAAATCCAGTACACGGCCATCGCACGCATCGCTGTGGCCATTGCAGGCTTAATCTTCTTCAGCT

ACGCACTAAGACCCAGCAGCACGTTGAGCTCTGGTGTGGAGAAGCTCGGCAACTTCACGCTCGGTCTGGT

ATCCTCTGTAGCTTTCCTCATTGGTGCTGTATGCTCCGCCACAGCAGGATACGTGAGCATGTGGGTCTCG

GSP2

GCCAGGTCCAATATCCGCGTGGCCAGTGCGGCTCGTGGATCTTATGGAGACGCGCTACTAGTCTGCTTCA

GAGGAGGAGCTTTCTCAGCTGTATTGGACATCACGCTCTGTGTCGGCGGAGTCAGTACACTGTATGTGAT

GSP1

GCTGTACATG

Figure 3.16 5' region of the *PnVPP* gene sequence showing the location of primers used in 5' RACE. Red boxes indicate the two potential translation start sites, located 285 bp apart. The three primers in grey boxes were used for reverse transcription (GSP1), first round PCR (GSP2) and second round PCR (GSP3).

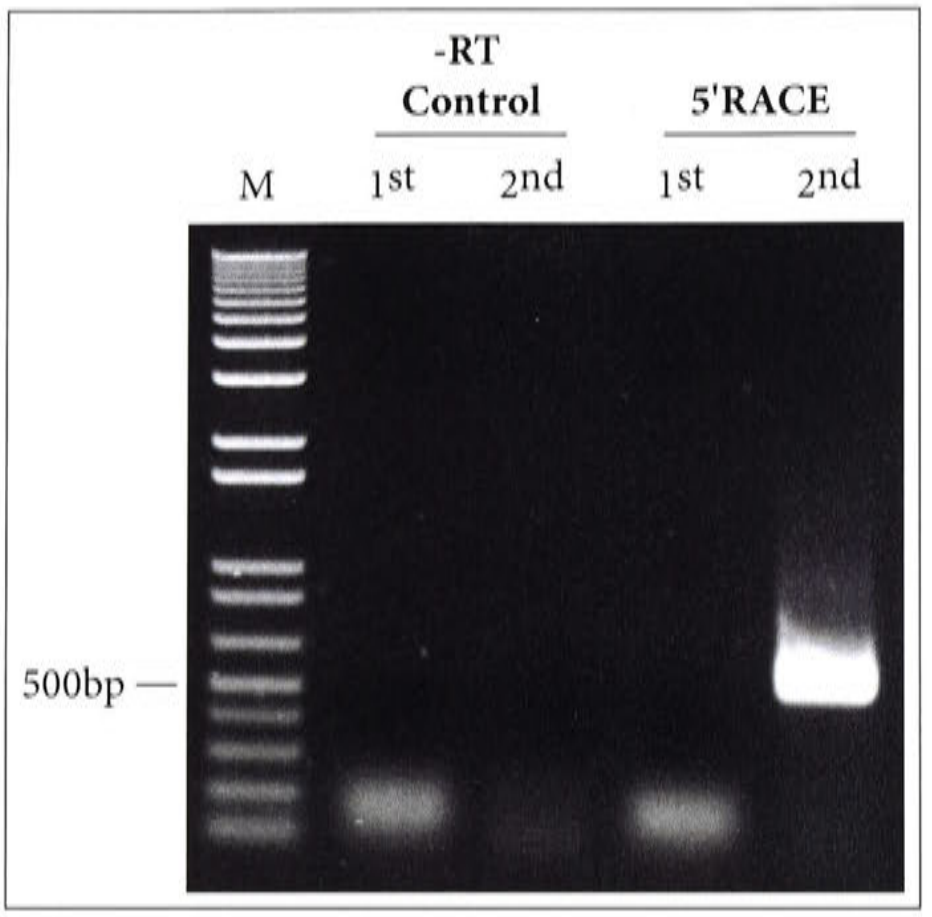


Figure 3.17 Isolation of 5'end of *PnVPP* transcript from *P. nicotianae* zoospore RNA after 5'RACE. The amplification of dA-tailed first strand cDNA using anchor specific primer 'Anchor1' and gene specific primer 'GSP2' is shown in lanes denoted '1st'. No visible products were observed. A second round PCR (lanes '2nd') with primers 'Anchor 2' and 'GSP3' resulted in a ~500 bp product as shown in the '5'RACE' labelled lane. The RT was omitted from the cDNA synthesis step in the '-RT Control' lanes, therefore no product was expected.

3.2.4 Isolation, subcloning and sequencing of a *PnVPP* genomic clone

The insert of cDNA clone DS311 was used to screen a genomic *P. nicotianae* BAC library according to protocols described in Section 2.2.8. The first three lanes of the blot, representing genomic DNA digested with restriction enzymes *Bam*HI, *Eco*RI and *Hin*dIII, indicate that *PnVPP* is a single copy gene in *P. nicotianae*. Five out of seven lanes of pooled BAC DNA contain the *PnVPP* gene (Figure 3.18). Two filters containing positive clones were screened with ^{32}P-labelled cDNA clone DS311. Both filters produced a single putative positive clone (Figure 3.19, 2G19 and 21K17). BAC DNA was isolated from both clones. Restriction digests with enzymes *Eco*RI, *Hin*dIII and *Bam*HI and subsequent Southern blot analysis using the DS311 probe showed that the clones were not identical (Figure 3.20). The ~6 kb *Bam*HI band from BAC clone 21K17 was subcloned into pBS vector and one clone was sequenced in both directions after confirmation of the correct inserts by Southern blot analysis (Figure 3.21).

The alignment of the genomic clone with cDNA clone DS311 revealed a single 70 bp intron with a 5' splice junction of GT/GGTAC and a 3' junction of GTA/G. Due to the

absence of a full length cDNA clone, it was initially not possible to determine the correct translation start codon in the genomic clone. Two potential ATG start codons positioned 285 bp apart were observed in the correct open reading frame (ORF). After removal of the intron, sequences encoded for either a 801 or 896 aa protein. Alignment of the full length cDNA information obtained through screening of the zoospore λ-library and 5' RACE (see 3.2.3) revealed the correct translation start site (Figure 3.22). Translation from the ATG present at nucleotide 728 of the genomic sequence encodes a 2,476 bp transcript resulting in a 801 aa protein with a predicted molecular weight of 84 kDa and a theoretical pI of 5.29.

The region upstream of the *PnVPP* gene showed two potential CAAT sequences (-140, -280), a CT-rich element from -44 to -23 as well as a TTATA sequence (- 237) which resembles the TATA/Pribnow box found in many eukaryotic and prokaryotic organisms. A sequence resembling the conserved motif GCYCATTYYNCAWTTTNYY found in the promoter region of many other oomycete genes (Pieterse *et al.*, 1994; McLeod *et al.*, 2004) was also observed in the 5' untranslated region (UTR) of the *PnVPP* gene starting at position -79.

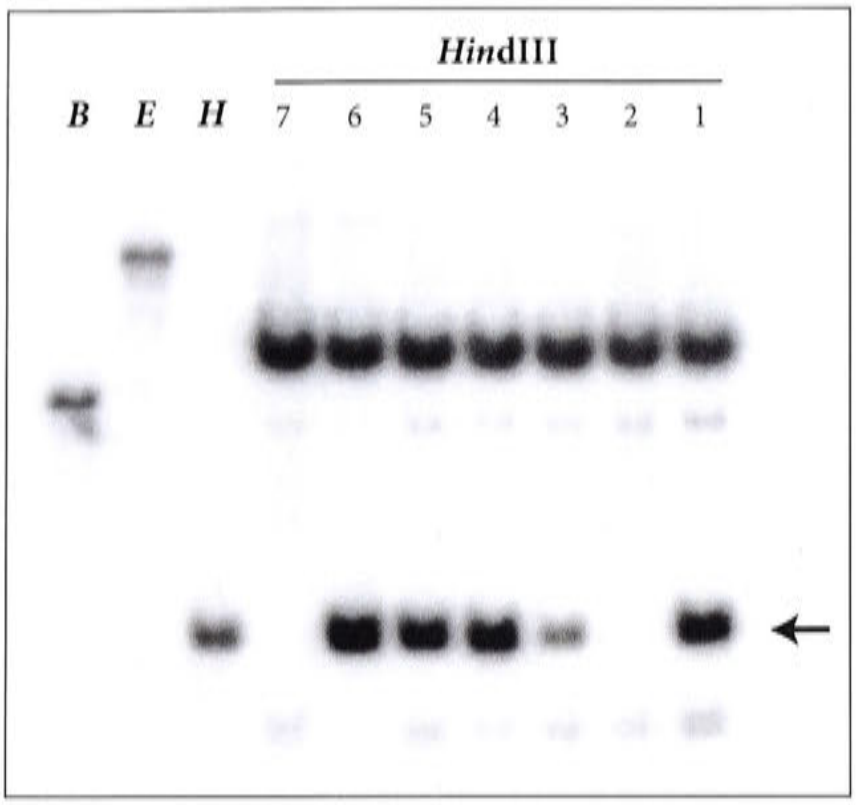


Figure 3.18 Screening of pooled BAC DNA from a *P. nicotianae* genomic library digested with *Hin*dIII (lanes 1-7: representing 1536 clones each) and genomic DNA of *P. nicotianae* digested with restriction enzymes *Bam*HI (*B*), *Eco*RI (*E*) and *Hin*dIII (*H*) and probed with ^{32}P-labeled cDNA clone DS311. The signal in *Hin*dIII digest lanes in the pooled BAC DNA is consistent with the band (arrow) in the *Hin*dIII restriction digested lane of *P. nicotianae* genomic DNA indicating that five out seven pooled BAC membranes contain putative *PnVPP* genomic clones.

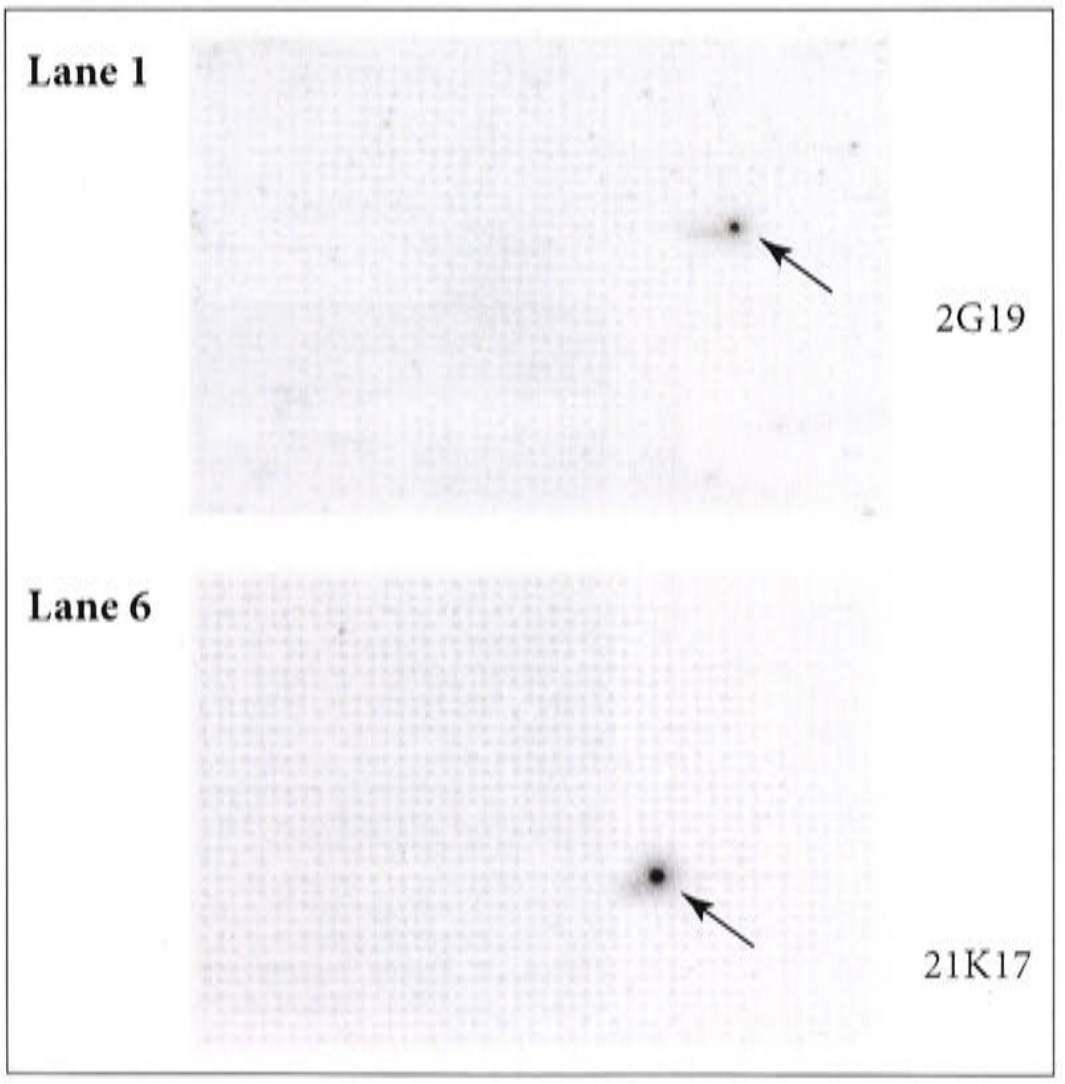


Figure 3.19 Two membranes from the BAC pool (lane 1 and 6) were hybridised with ^{32}P-labelled cDNA cloned DS311. Both filters produced a single putative positive clone each (arrows, 2G19 and 21K17).

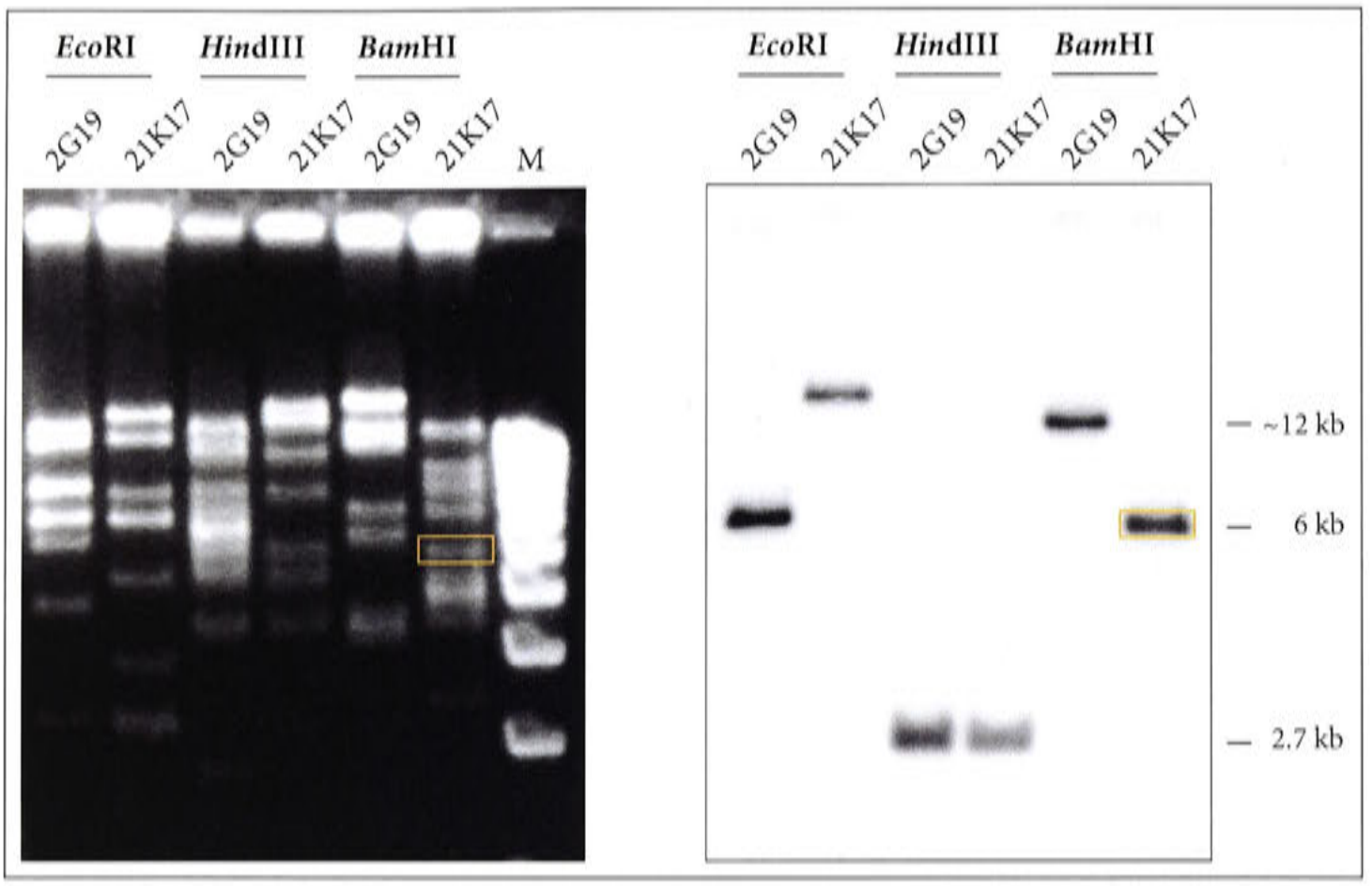


Figure 3.20 Restriction digests of two putative *PnVPP P. nicotianae* BAC genomic clones with enzymes *Eco*RI, *Hin*dIII and *Bam*HI in order to identify potential restriction sites for cloning the genomic clone into pBS vector. The ~6 kb *Bam*HI band (orange box) from genomic clone 21K17 was selected for subcloning in pBS to enable sequencing of the complete genomic clone.

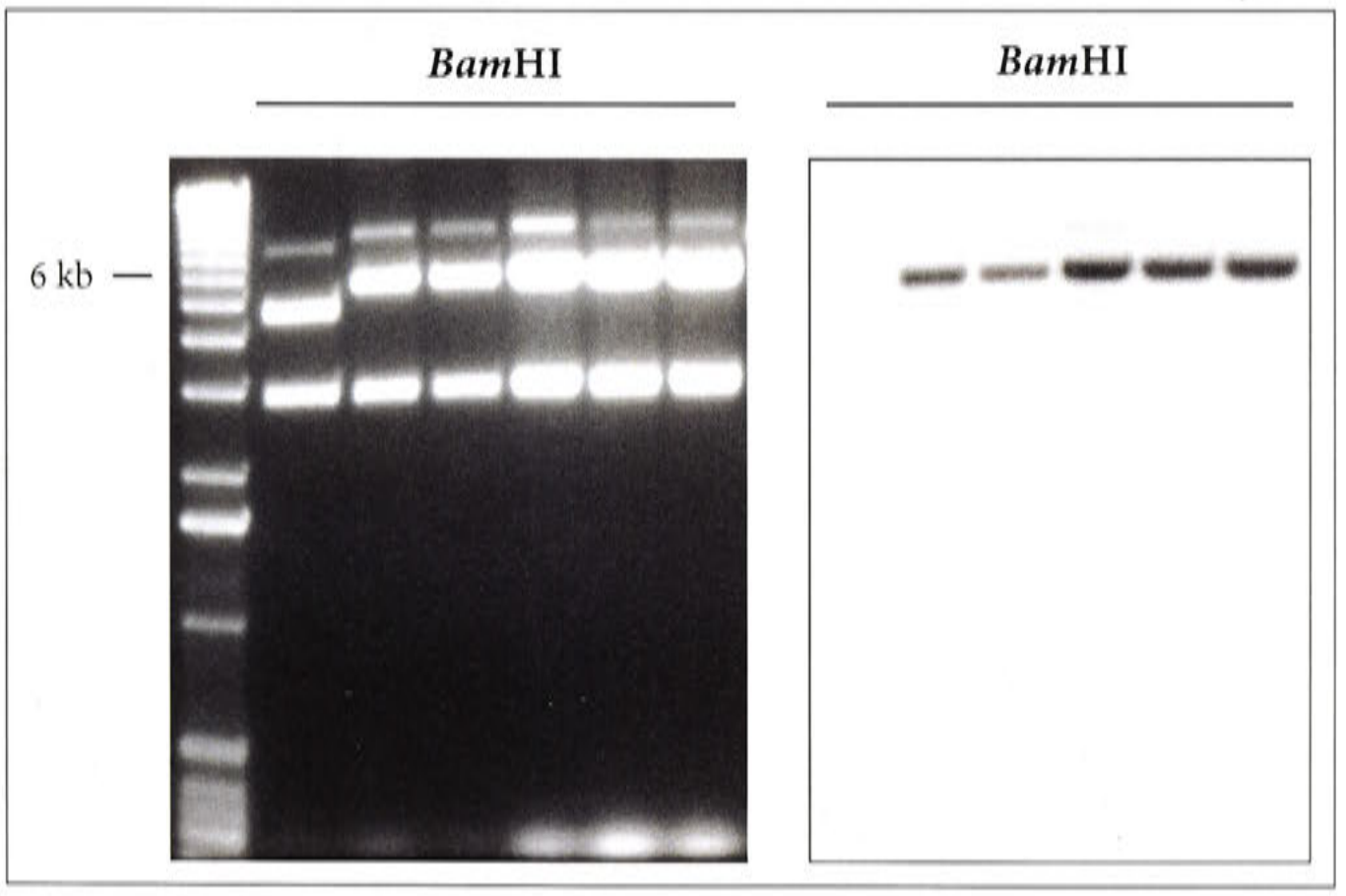


Figure 3.21 *Bam*HI restriction digests of six *PnVPP* genomic subclones obtained from pBS + *PnVPP* 21K17 ~6 kb *Bam*HI ligation reactions. The *Bam*HI digests of six subclones were probed with the ^{32}P-labelled cDNA clone DS311. The picture on the left shows the EtBr stained gel, the picture on the right depicts the corresponding hybridisation results.

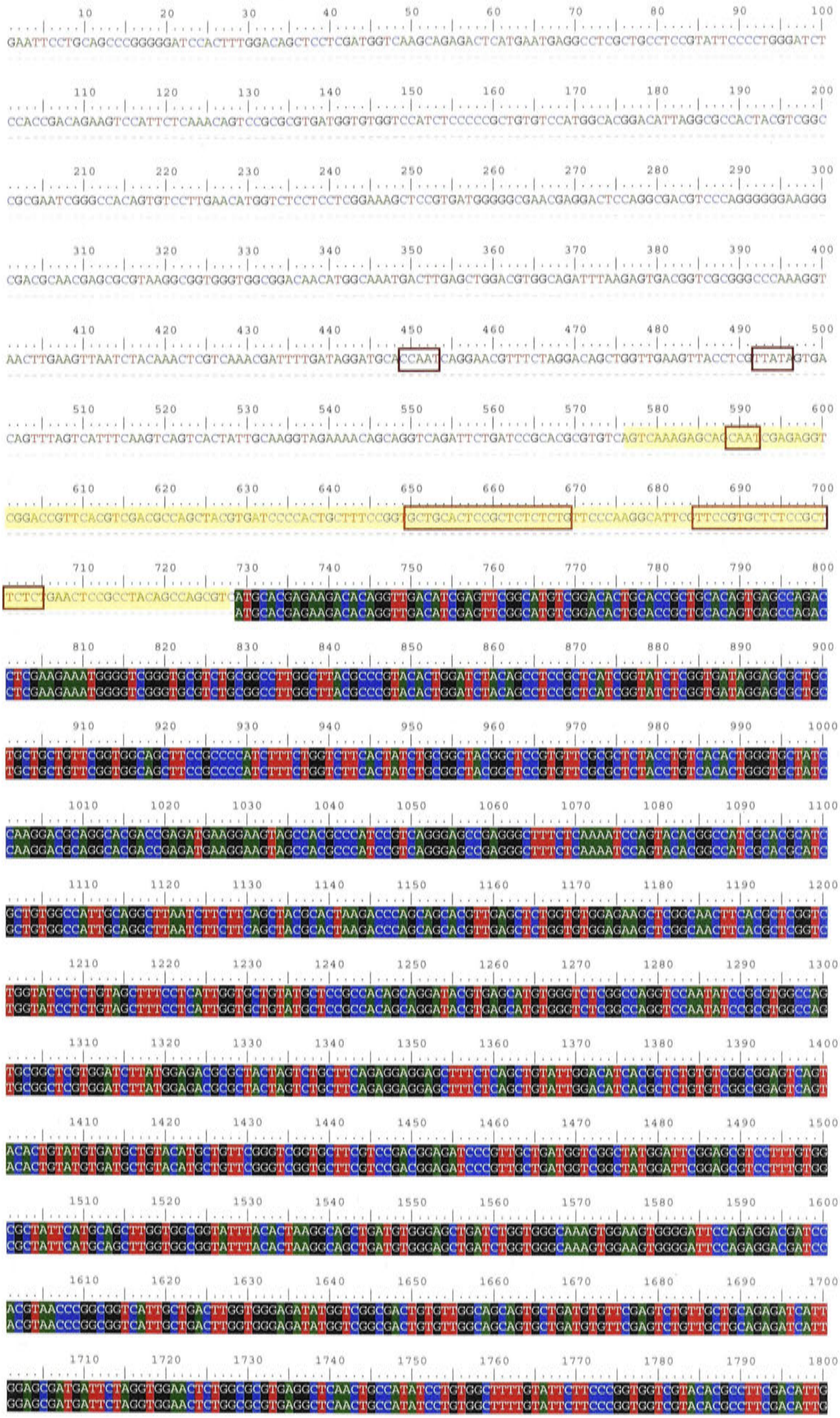

10 20 30 40 50 60 70 80 90 100
GAATTCCTGCAGCCCGGGGGATCCACTTTGGACAGCTCCTCGATGGTCAAGCAGAGACTCATGAATGAGGCCTCGCTGCCTCCGTATTCCCCTGGGATCT
110 120 130 140 150 160 170 180 190 200
CCACCGACAGAAGTCCATTCTCAAACAGTCCGCGCGTGATGGTGTGGTCCATCTCCCCCGCTGTGTCCATGGCACGGACATTAGGCGCCACTACGTCGGC
210 220 230 240 250 260 270 280 290 300
CGCGAATCGGGCCACAGTGTCCTTGAACATGGTCTCCTCCTCGGAAAGCTCCGTGATGGGGGCGAACGAGGACTCCAGGCGACGTCCCAGGGGGGAAGGG
310 320 330 340 350 360 370 380 390 400
CGACGCAACGAGCGCGTAAGGCGGTGGGTGGCGGACAACATGGCAAATGACTTGAGCTGGACGTGGCAGATTTAAGAGTGACGGTCGCGGGCCCAAAGGT
410 420 430 440 450 460 470 480 490 500
AACTTGAAGTTAATCTACAAACTCGTCAAACGATTTTGATAGGATGCACCAATCAGGAACGTTTCTAGGACAGCTGGTTGAAGTTACCTCGTTATAGTGA
510 520 530 540 550 560 570 580 590 600
CAGTTTAGTCATTTCAAGTCAGTCACTATTGCAAGGTAGAAAACAGCAGGTCAGATTCTGATCCGCACGCGTGTCAGTCAAAGAGCAGCAATCGAGAGGT
610 620 630 640 650 660 670 680 690 700
CGGACCGTTCACGTCGACGCCAGCTACGTGATCCCCACTGCTTTCCGGTGCTGCACTCCGCTCTCTCTGTTCCCAAGGCATTCCTTCCGTGCTCTCCGCT
710 720 730 740 750 760 770 780 790 800
TCTCTGAACTCCGCCTACAGCCAGCGTCATGCACGAGAAGACACAGGTTGACATCGAGTTCGGCATGTCGGACACTGCACCGCTGCACAGTGAGCCAGAC
ATGCACGAGAAGACACAGGTTGACATCGAGTTCGGCATGTCGGACACTGCACCGCTGCACAGTGAGCCAGAC
810 820 830 840 850 860 870 880 890 900
CTCGAAGAAATGGGGGTCGGGTGCGTCTGCGGCCTTGGCTTACGCCCGTACACTGGATCTACAGCCTCCGCTCATCGGTATCTCGGTGATAGGAGCGCTGC
CTCGAAGAAATGGGGGTCGGGTGCGTCTGCGGCCTTGGCTTACGCCCGTACACTGGATCTACAGCCTCCGCTCATCGGTATCTCGGTGATAGGAGCGCTGC
910 920 930 940 950 960 970 980 990 1000
TGCTGCTGTTCGGTGGCAGCTTCCGCCCCATCTTTCTGGTCTTCACTATCTGCGGCTACGGCTCCGTGTTCGCGCTCTACCTGTCACACTGGGTGCTATC
TGCTGCTGTTCGGTGGCAGCTTCCGCCCCATCTTTCTGGTCTTCACTATCTGCGGCTACGGCTCCGTGTTCGCGCTCTACCTGTCACACTGGGTGCTATC
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
CAAGGACGCAGGCACGACCGAGATGAAGGAAGTAGCCACGCCCATCCGTCAGGGAGCCGAGGGCTTTCTCAAAATCCAGTACACGGCCATCGCACGCATC
CAAGGACGCAGGCACGACCGAGATGAAGGAAGTAGCCACGCCCATCCGTCAGGGAGCCGAGGGCTTTCTCAAAATCCAGTACACGGCCATCGCACGCATC
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
GCTGTGGCCATTGCAGGCTTAATCTTCTTCAGCTACGCACTAAGACCCAGCAGCACGTTGAGCTCTGGTGTGCAGAAGCTCGGCAACTTCACGCTCGGTC
GCTGTGGCCATTGCAGGCTTAATCTTCTTCAGCTACGCACTAAGACCCAGCAGCACGTTGAGCTCTGGTGTGGAGAAGCTCGGCAACTTCACGCTCGGTC
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
TGGTATCCTCTGTAGCTTTCCTCATTGGTGCTGTATGCTCCGCCACAGCAGGATACGTGAGCATGTGGGTCTCGGCCAGGTCCAATATCCGCGTGGCCAG
TGGTATCCTCTGTAGCTTTCCTCATTGGTGCTGTATGCTCCGCCACAGCAGGATACGTGAGCATGTGGGTCTCGGCCAGGTCCAATATCCGCGTGGCCAG
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
TGCGGCTCGTGGATCTTATGGAGACGCGCTACTAGTCTGCTTCAGAGGAGGAGCTTTCTCAGCTGTATTGGACATCACGCTCTGTGTCGGCGGAGTCAGT
TGCGGCTCGTGGATCTTATGGAGACGCGCTACTAGTCTGCTTCAGAGGAGGAGCTTTCTCAGCTGTATTGGACATCACGCTCTGTGTCGGCGGAGTCAGT
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
ACACTGTATGTGATGCTGTACATGCTGTTCGGGTCGGTGCTTCGTCCGACGGAGATCCCGTTGCTGATGGTCGGCTATGGATTCGGAGCGTCCTTTGTGG
ACACTGTATGTGATGCTGTACATGCTGTTCGGGTCGGTGCTTCGTCCGACGGAGATCCCGTTGCTGATGGTCGGCTATGGATTCGGAGCGTCCTTTGTGG
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
CGCTATTCATGCAGCTTGGTGGCGGTATTTACACTAAGGCAGCTGATGTGGGAGCTGATCTGGTGGGCAAAGTGGAAGTGGGGATTCCAGAGGACGATCC
CGCTATTCATGCAGCTTGGTGGCGGTATTTACACTAAGGCAGCTGATGTGGGAGCTGATCTGGTGGGCAAAGTGGAAGTGGGGATTCCAGAGGACGATCC
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
ACGTAACCCGGCGGTCATTGCTGACTTGGTGGGAGATATGGTCGGCGACTGTGTTGGCAGCAGTGCTGATGTGTTCGAGTCTGTTGCTGCAGAGATCATT
ACGTAACCCGGCGGTCATTGCTGACTTGGTGGGAGATATGGTCGGCGACTGTGTTGGCAGCAGTGCTGATGTGTTCGAGTCTGTTGCTGCAGAGATCATT
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
GGAGCGATGATTCTAGGTGGAACTCTGGCGCGTGAGGCTCAACTGCCATATCCTGTGGCTTTTGTATTCTTCCCGGTGGTCGTACACGCCTTCGACATTG
GGAGCGATGATTCTAGGTGGAACTCTGGCGCGTGAGGCTCAACTGCCATATCCTGTGGCTTTTGTATTCTTCCCGGTGGTCGTACACGCCTTCGACATTG

Figure 3.22 Alignment of *PnVPP* genomic sequence contig with the complete cDNA sequence obtained from two cDNA clones and 5' RACE. The 2406 bp boxed region denotes the *PnVPP* cDNA sequence aligned with the *PnVPP* genomic BAC library clone. There is perfect alignment between the cDNA (starting at ATG in shades) and genomic sequence of *PnVPP* except for the single 70 bp intron (2822-2891) present in the genomic clone of *PnVPP*. Red boxes denote the potential promoter binding regions (CAAT-box, TATA-box CT-rich regions and 19 bp motif). The yellow shaded area displays the 5' UTR of the mRNA as inferred from 5' RACE experiments.

3.2.5 PnVPP protein, homologies and conserved domains

The full length amino acid sequence was subjected to another protein-protein BLAST search to ascertain the best homology hit. Interestingly, the search with the 801 aa sequence produced a different result compared to the BLAST search that used the cDNA clone DS311 described in Section 3.2.1. The most similar protein to PnVPP was again an *A. thaliana* vacuolar-type H^+-translocating inorganic pyrophosphatase, however, not the AVP1 protein reported above but the Type-II V-PPase AVP2 (GenBank accession number AAF31163).

Application of several transmembrane region prediction programs and a hydropathy plot predicted a structure containing 15-17 transmembrane helices (Figure 3.23, Figure 3.24 – overlay, Table 3.4). All of the sequence motifs demonstrated to be necessary for catalysis by V-PPases were conserved in PnVPP. Furthermore, features of the polypeptide characteristic for Type-II V-PPases were present. These include the following (Figure 3.24): a) the putative binding site and catalytic domain motif Dx_7KxE located in the cytosolic loop *f*, b) the two acidic motifs Dx_3Dx_3D situated in loops *f* and *p*, respectively, proximal to the TMDs, c) the Type I/Type II distinguishing motif in loop j – [DE]YYTS in Type I and [KQR]YYTS in Type II V-PPases, respectively, d) the conserved GNxx[K/A]Ax[TA/G] motif, also used for distinguishing Type I from Type II V-PPases and conferring the requirement for K^+ (Figure 3.25), and e) the GYG motif located in TMD6 (residues 249-251), which potentially plays a significant role in maintaining structure and function of V-PPases (Van *et al.*, 2005). Furthermore, some important residues have also been detected: a) C666 corresponding to C668 of AVP2, whose alkylation by maleimides irreversibly abolishes catalytic activity (Zhen *et al.*, 1994b; Kim *et al.*, 1995), and b) one of the two acidic residues, E322 (corresponding to E323 in AVP2), which contribute to the susceptibility of the enzyme to inhibition by DCCD (Zhen *et al.*, 1997). Interestingly, the second acidic residue D532 in AVP2 has been replaced with a neutral asparagine (N530) in PnVPP.

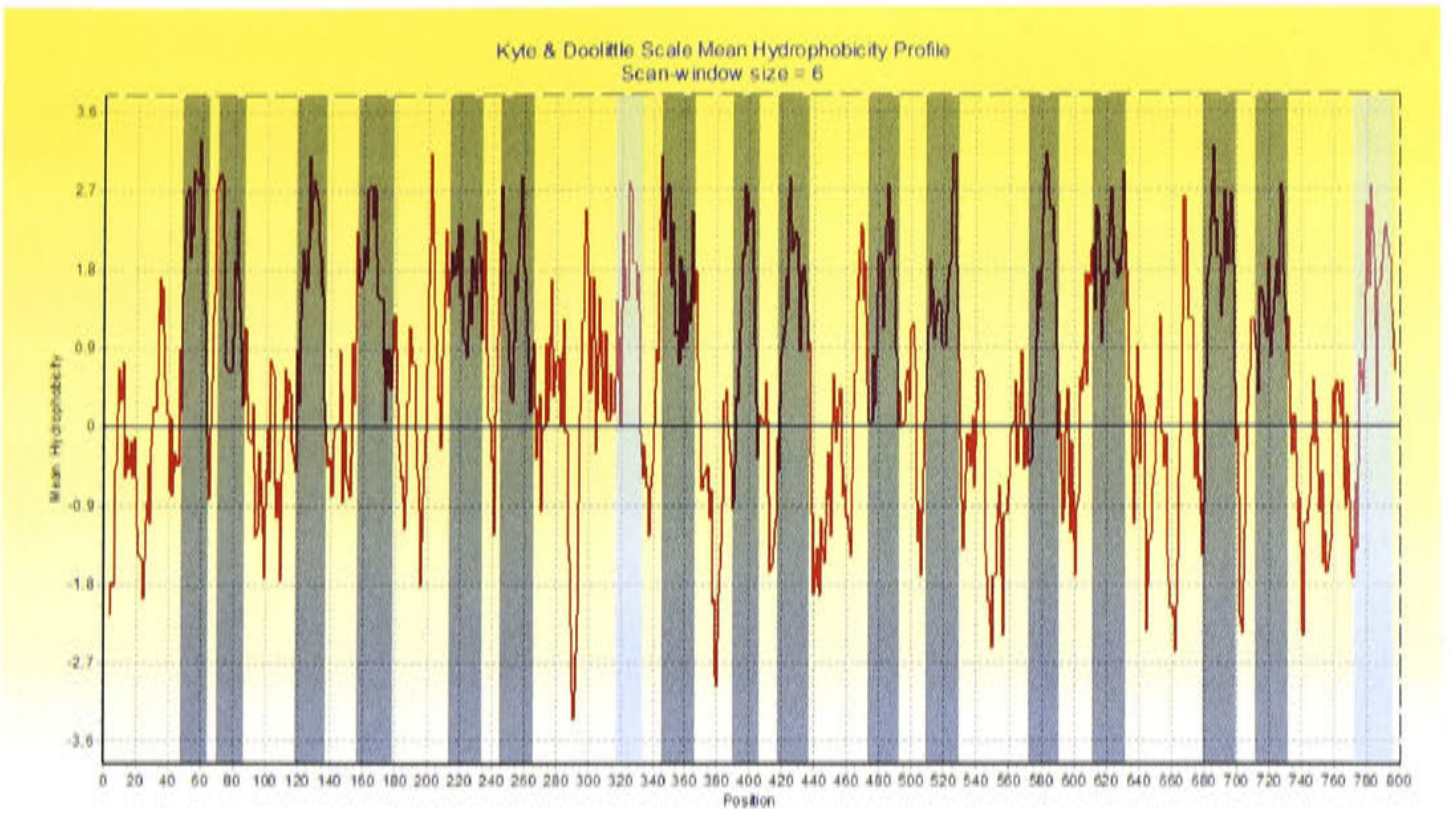


Figure 3.23 The Kyle and Doolittle hydropathy profile shows the hydrophobic and hydrophilic regions of the PnVPP amino acid sequence. The plotted regions below '0' (values on the Y-axis) constitute hydrophilic amino acids in the sequence while the plotted regions above '0' (values on the Y-axis) constitute hydrophobic amino acids in the sequence. Grey areas show predicted transmembrane regions as predicted by several programs (Table 3.4).

Table 3.4 Prediction of transmembrane regions in PnVPP by five prediction programs. (CoPreTHi - Promponas *et al.*, 1999, SMART - Schultz *et al.*, 2000, PHD - Rost *et al.*, 1994, Hmmtop - Tusnady and Simon, 2001, SPLI - Juretic *et al.*, 2002)

	CoPreTHi		SMART		PHD		Hmmtop		SPLIT	
	from	to	from	to	from	to	from	to	from	to
1	47	65	47	64	43	60	46	63	46	64
2	68	85	68	90	65	82	68	90	68	88
3	118	138	116	138	118	135	121	138	115	138
4	156	177	158	180	158	176	159	183	154	178
5	216	234	209	231	210	227	214	238	208	236
6	243	263	246	268	247	264	243	267	243	266
7	2 of 4 *				319	336	315	334	315	333
8	344	366	343	365	347	364	343	367	340	368
9	389	409	388	410	389	406	388	405	390	406
10	417	436	417	436	421	438	418	435	418	441
11	475	491	478	500	471	489	466	488	464	494
12	509	529	507	529	509	526	509	528	509	531
13	572	590	568	590	572	589	573	590	574	591
14	611	631	611	633	612	629	611	633	605	635
15	680	698	678	700	680	697	680	699	679	699
16	712	732	707	729	713	730	712	731	709	733
17	2 of 4 *		776	798	780	797	781	798	776	796

* indicates that 2 out 4 programs used by CoPreTHi predicted a transmembrane loop at this position

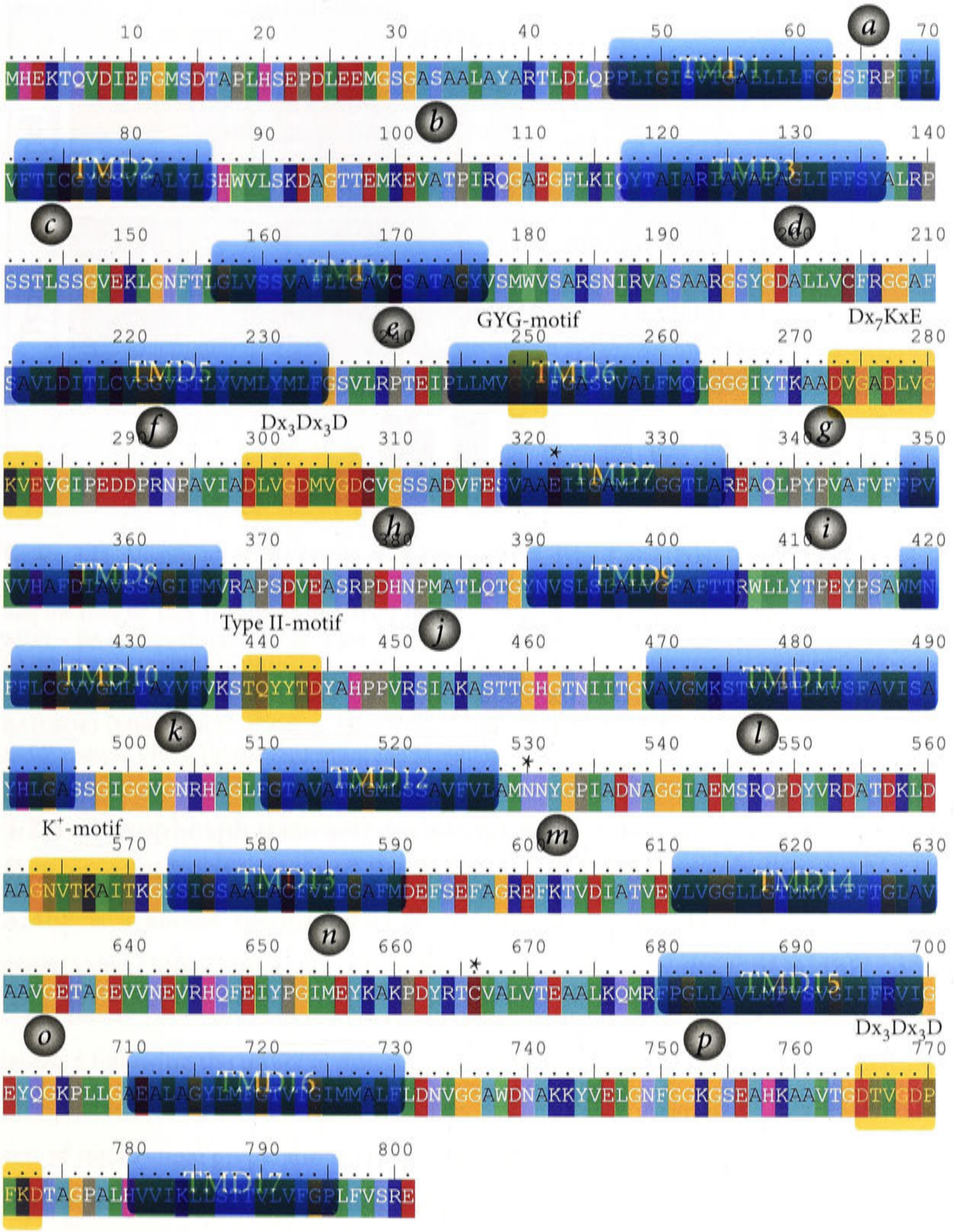


Figure 3.24 PnVPP amino acid sequence with sequence motifs. Orange boxes indicate conserved domains, asterisks denote amino acid residues of interest. The overlay shows predicted TMD and corresponding loops. See text for details.

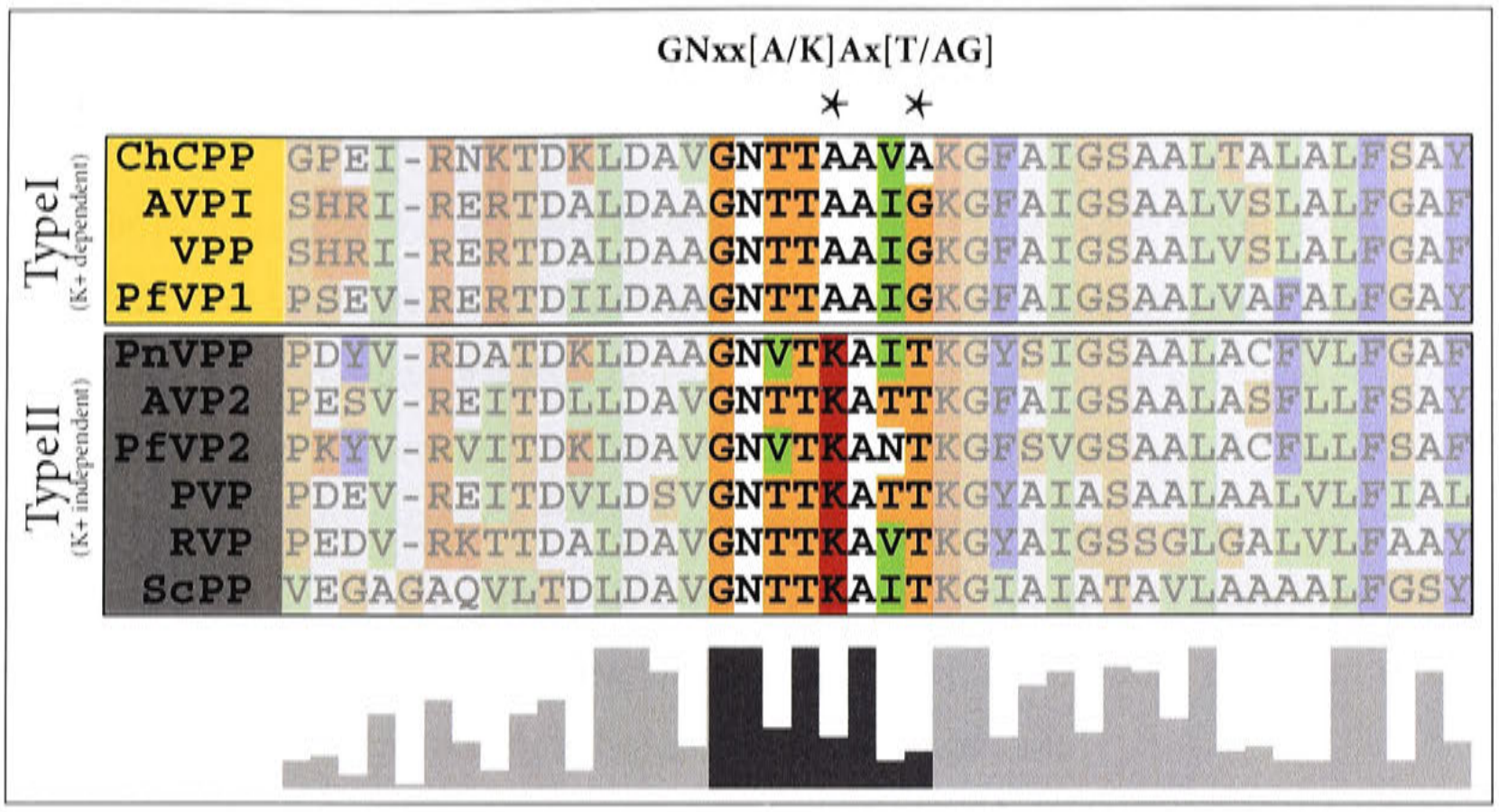


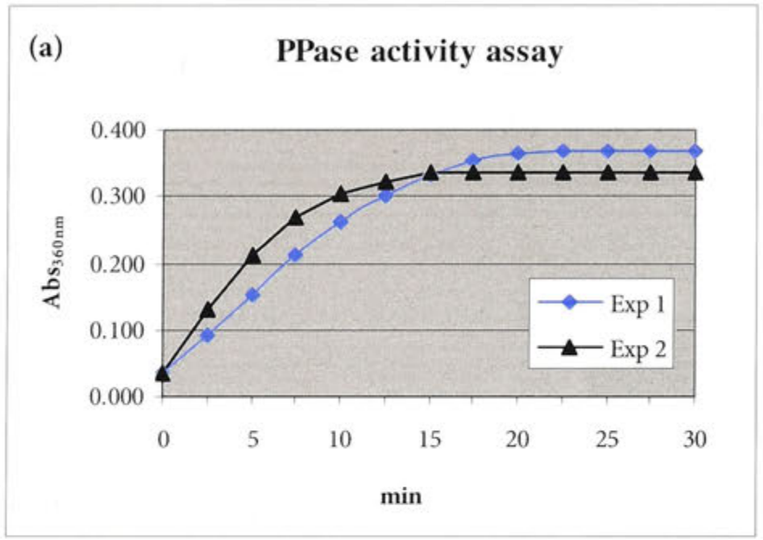
Figure 3.25 Alignment of the amino acid sequences of Type I and Type II V-PPases around the conserved motif GNxx[A/K]Ax[T/AG]. Asterisks denote amino acid residues of interest. PnVPP falls into the class of Type II V-PPase displaying the representative residues K and T, respectively. *A. thaliana*: AVP1 - BAA32210 (Sarafian *et al.*, 1992) and AVP2 – AAF31163 (Drozdowicz *et al.*, 2000), *P. faliparum*: PfVP1 – AAD17215 and PfVP2 - AAG21366 (McIntosh *et al.*, 2001), *P. aerophilum*: PVP – AAF01029 (Drozdowicz *et al.*, 1999), *R. rubrum*: RVP – AAC38615 (Baltscheffsky *et al.*, 1998), *S. coelicolor*: ScPP - BAD36743 (Mimura *et al.*, 2004), *C. hydrogenoformans*: ChCPP - ABB14908 (Belogurov and Lahti, 2002) and *V. radiata*: VPP – T07801 (Nakanishi and Maeshima, 1998).

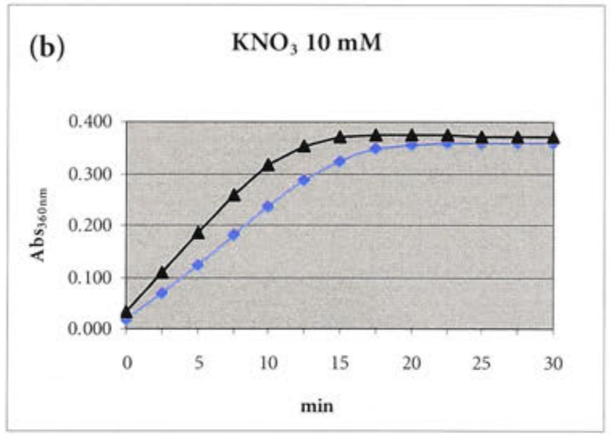
3.2.6 Pyrophosphatase activity assay on microsomal preparations of *P. nicotianae* zoospores and the influence of inhibitor substance AMBP

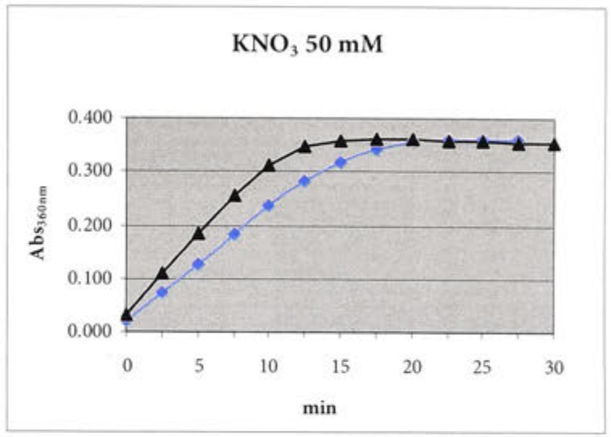
AMBP is a known specific inhibitor of V-PPase. A series of *in vitro* and *in vivo* experiments was carried out in order to determine the effect of AMBP on the *P. nicotianae* V-PPase. Microsomal preparations from *P. nicotianae* zoospores were isolated following the protocol described in Section 2.6.2 and subjected to the EnzCheck Phosphatase Assay Kit (Molecular Probes, Sydney, NSW, Australia). Two independent sets of microsomal preparations were used in the experiment. Controls as well as treated samples were analysed in triplicates. Samples were treated with a series of concentrations of either AMBP or KNO_3, a known inhibitor of the V-ATPase (Mitchell and Hardham, 1999, Figure 3.26). Activity reduction was determined by calculation of relative P_i release between 2.5 min and 7.5 min (Δ 2.5-7.5) using a standard curve and comparison of treated sample values to control samples. Table 3.5 shows the reduction of pyrophosphatase activity by 40-57% when using a concentration of 30 µM AMBP. Presence of KNO_3 had no significant affect on pyrophosphatase activity in microsomal

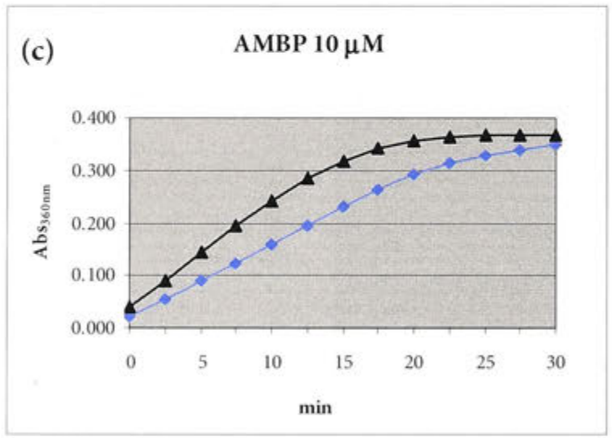
preparations of *P. nicotianae* zoospores. Higher concentrations of AMBP did not lead to a further reduction in activity (data not shown).

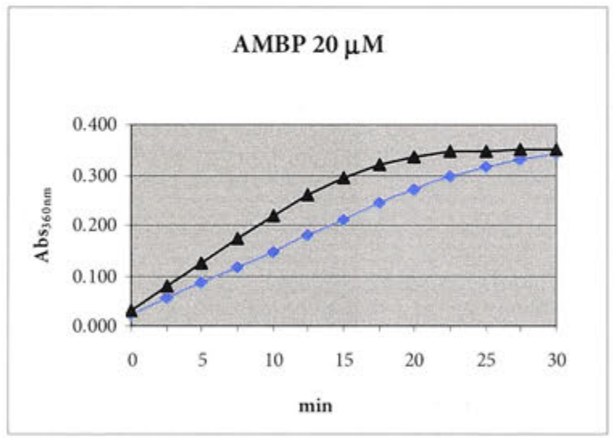
In order to determine the influence of AMBP on the V-PPase in living zoospores and its potential effect on the WEV video analysis was carried out. Zoospores were treated with a range of concentrations of KNO_3 (2 – 20 mM) and AMBP (2 – 10 μM), and zoospores whose movement was constrained by surrounding hyphae were captured on video for approximately 1 min (Figure 3.27). Video recordings were analysed and the WEV cycle time was measured. Only a small dataset was obtained. Figure 3.28 shows that KNO_3 slows the WEV cycle down quite significantly, however, AMBP does not seem to have any impact on WEV function or zoospore behaviour in general.











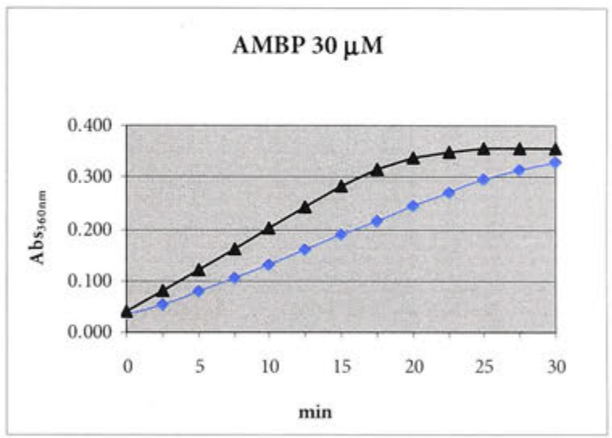


Figure 3.26 Effects of KNO_3 and AMBP on kinetics of pyrophosphatase activity in microsomal preparations from *P. nicotianae* zoospores. Graphs represent two independent sets of experiments. PP_i was used as a substarte in all experiments at a concentration of 300 µM. (a) Pyrophosphatase activity in untreated samples, (b) addition of 10 mM or 50 mM KNO_3 to the samples prior to activity measurement does not influence enzyme activity, and (c) AMBP was added to the assay in three different concentrations (10 µM, 20 µM and 30 µM). A reduction in activity was observed at all three AMBP concentrations with 30 µM showing the greatest decrease.

Table 3.5 Calculation of pyrophosphatase activity reduction in microsomal membrane preparations from *P. nicotianae* zoospores. Data shown are from two independent sets of experiments.

	Δ2.5-7.5 min (P_i in nmol)		**μmol PP_i/min/ mg protein**		**Activity reduction**	
	Exp. 1	*Exp. 2*	*Exp. 1*	*Exp. 2*	*Exp. 1*	*Exp. 2*
No inhibitor	4.90	5.52	0.097	0.110		
AMBP 10 μM	2.75	4.18	0.055	0.084	**43%**	**23%**
AMBP 20 μM	2.40	3.86	0.048	0.077	**51%**	**29%**
AMBP 30 μM	2.05	3.28	0.041	0.066	**57%**	**40%**
KNO_3 10 mM	4.56	6.04	0.091	0.121		
KNO_3 50 mM	4.40	5.92	0.088	0.119		

Figure 3.27 Video analysis of *P. nicotianae* zoospores. Image of a trapped *P. nicotianae* zoospore captured during video analysis.

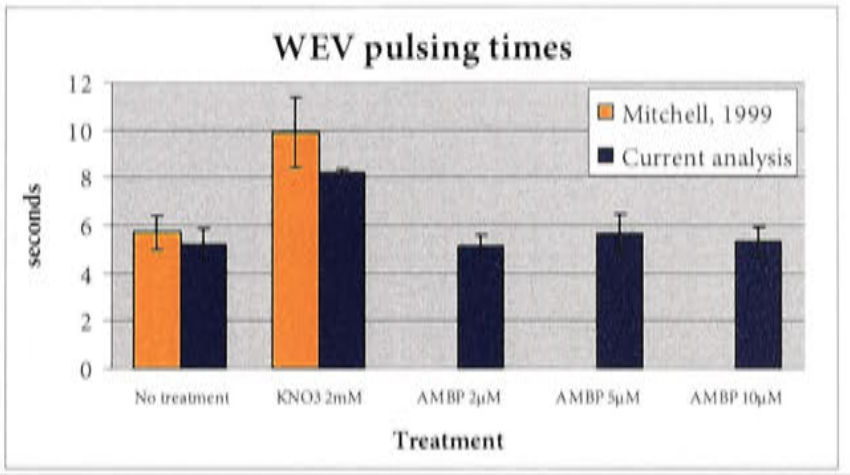
Mitchell and Hardham, 1999

	Mean cycle time	n	Increase in cycle time
No treatment	5.70 ± 0.71s	111	
2 mM KNO_3	9.92 ± 1.46s	32	174%

Current analysis

No treatment	5.21 ± 0.64s	10	
2 mM KNO_3	8.21 ± 0.15s	2	157%
2 µM AMBP	5.13 ± 0.45s	7	98%
5 µM AMBP	5.65 ± 0.79s	9	108%
10 µM AMBP	5.31 ± 0.64s	10	102%

a)



b)

Figure 3.28 Results of video analysis of *P. nicotianae* zoospore WEV cycle times after treatment with different agents. (a) Reference data obtained from Mitchell and Hardham, (1999) and experimental data of *in vivo* video analysis of *P. nicotianae* zoospores. The mean cycle time of a number of zoospores (n) and the calculated change of cycle times in percent compared to untreated samples is shown, (b) graphic display of data obtained. No reduction in the WEV mean cycle time was observed after treatment of *P. nicotianae* zoospores with AMBP.

3.2.7 Overexpression of a PnVPP fusion protein for the generation of antibodies

In order to generate antibodies against PnVPP it was attempted to overexpress part of the protein using the pETBlue-2 vector system (Novagen, Madison, WI, USA). Predicted transmembrane regions (Figure 3.23, Table 3.4), as well as the fact that only a partial cDNA clone was available for PCR amplification of segments were taken into account when regions where selected. Two regions were chosen for overexpression: a 153 bp fragment ranging from amino acid positions 730 – 780 resulting in a 6.6 kDa protein and a 453 bp fragment from positions 532 – 682 resulting in a 17.7 kDa protein (Figure 3.29). Regions were PCR amplified from *PnVPP* cDNA clone E-20 and subcloned into the pETBlue-2 vector. Colonies obtained were tested for correct insertion of fragments by colony PCR and selected clones were fully sequenced (data not shown). For overexpression, plasmids were transformed into the *E. coli* cell line XL1-Blue or Tuner™(DE3)pLacI (Novagen) and subjected to IPTG induction. Induction time as well as IPTG concentration was varied over a set of experiments to achieve protein expression, however, no expression was observed (Figure 3.30). Further attempts with bigger fragments (744 bp and 1244 bp) as well as using a different overexpression system (QIAexpress, vector pQE30Xa) failed to produce the target protein (data not shown).

153 bp fragment

Amino acid positions 730-780

ATTCCTCGACAACGTGGGCGGCGCGTGGGATAACGCCAAGAAATACGTCGAGTTGGGCAACTTCGGAG
GCAAAGGCAGCGAGGCGCACAAGGCTGCTGTCACTGGCGACACAGTGGGTGATCCGTTCAAGGATACGG
CTGGGCCTGCGCTGCACCTCGAGAATT

FLDNVGGAWDNAKKYVELGNFGGKGSEAHKAAVTGDTVGDPFKDTAGPALH

453 bp fragment

Amino acid positions 532-682

ATACGGCCCCATCGCTGATAACGCTGGAGGTATCGCCGAGATGAGTCGTCAGCCGGACTACGTGCGTG
ATGCTACGGACAAGCTGGATGCTGCTGGCAATGTGACAAAGGCTATTACTAAGGGATATTCGATCGGCT
CGGCTGCTCTTGCTTGCTTCGTGCTGTTCGGCGCTTTCATGGACGAGTTCTCAGAGTTCGCTGGTCGCG
AGTTTAAGACGGTGGATATTGCTACGGTGGAGGTCCTGGTTGGCGGTCTGCTGGGTACTATGATGGTCT
TCTTCTTTACTGGACTTGCTGTTGCTGCTGTGGGCGAGACCGCTGGTGAGGTTGTCAATGAAGTGCGCC
ACCAGTTTGAGATTTACCCGGGAATTATGGAGTACAAGGCGAAGCCCGACTACCGCACTTGCGTCGCGC
TTGTCACCGAGGCTGCATTGAAGCAGATGCGTTTCCCAGGGCTCGAGAAT

YGPIADNAGGIAEMSRQPDYVRDATDKLDAAGNVTKAITKGYSIGSAALACFVLFGAFMDEFSEFAGRE
FKTVDIATVEVLVGGLLGTMMVFFFTGLAVAAVGETAGEVVNEVRHQFEIYPGIMEYKAKPDYRTCVAL
VTEAALKQMRFPG

Figure 3.29 DNA sequences of selected regions for overexpression of a partial PnVPP protein. Both fragments were cloned into vector pETBlue-2 utilising ORF 3. Modified primers for PCR amplification are shown as boxed areas. The nucleic acid depicted in green was added to the primer sequence to achieve cloning in the correct reading frame. The blue area shows the incorporated restriction site *Xho*I.

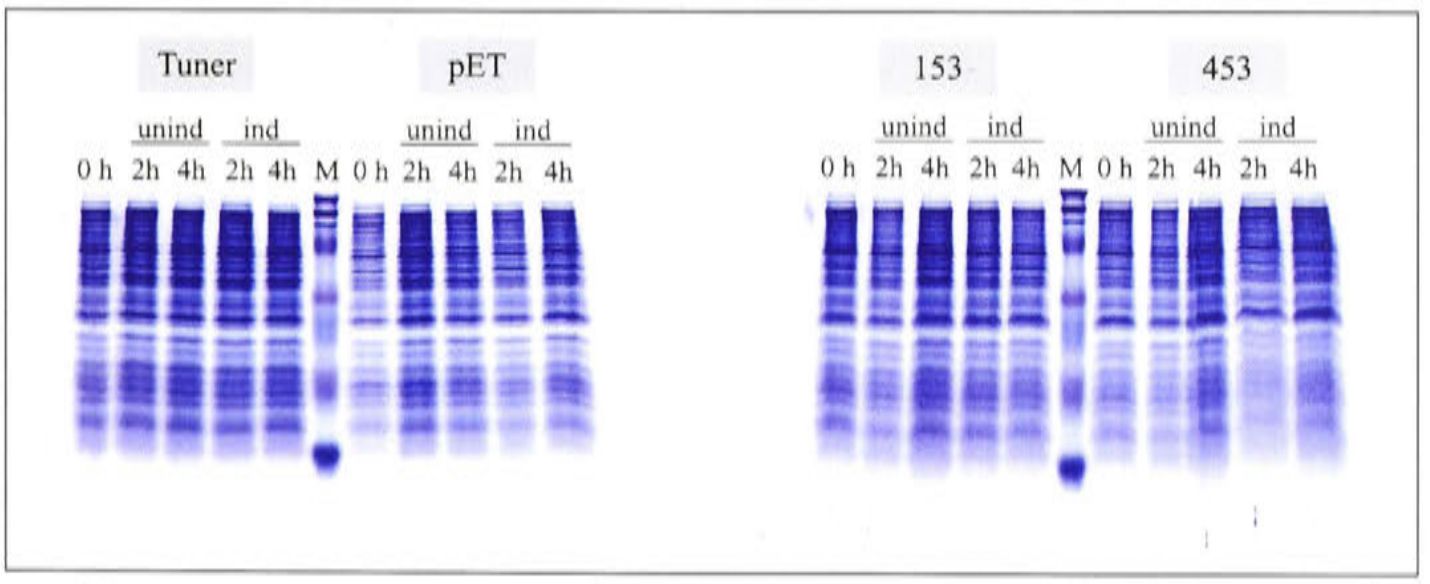


Figure 3.30 SDS-PAGE analysis followed by Coomassie blue staining of total cell protein after induction with 1 mM IPTG for the assessment of expression of the target protein. Cells were induced with IPTG for the indicated time and total cell protein was isolated. Uninduced cultures served as controls. (a) The pETBlue-2 compatible expression host strains Tuner™(DE3)pLacI alone (left of molecular weight marker – M) and transformed with an empty pETBlue-2 vector (right) was analysed, (b) cell lines transformed with the expression vector containing cassettes for the 153 bp (~6.6 kDa) and 453 bp (~17.7 kDa) fragments were analysed. No overexpression of protein was observed.

3.2.8 Peptide design for the production of antibodies

As the overexpression of the PnVPP protein was unsuccessful, peptides for the generation of antibodies were designed. Potential antigenicity sites (predicted by WebAngis, Table 3.6), hydrophobicity plots and predicted transmembrane loops (see above) as well as secondary structures, potential post-translation modification sites and required conjugation methods were considered when designing the peptides. The two peptide sequences chosen for immunisation are shown in Figure 3.31 and Figure 3.32.

Table 3.6 Predicted antigenic sites of generated peptides (WebAngis). Max score position at "*"

Score 1.151 length 19 at residues 269-287 * TKAADVGADLVGKVEVGIP | | 269 287	⇨	Inside
Score 1.074 length 8 at residues 758-765 * HKAAVTGD | | 758 765	⇨	Outside

Pep-PP1

Residues: DEK have acidic or basic side chains (underlined)
Residues: GAVLC have hydrophobic side chains **(bold)**
Residue: T has uncharged-polar side chains (*in italics*)

towards N terminus ◀ C *T* K **A A** D **V G A** D **L V G** K **V** E ▶ towards C terminus

Peptide analyses: 63% of the residues are hydrophobic, 31% are charged, and 6% are non-charged.

Pep-PP2

Residues: DKH have acidic or basic side chains (underlined)
Residues: GAVCPF have hydrophobic side chains **(bold)**
Residue: T has uncharged-polar side chains (*in italics*)

towards N terminus ◀ H K **A A V** *T* **G** D *T* **V G** D **P F** K C ▶ towards C terminus

Peptide analyses: 50% residues are hydrophobic, 38% are charged, and 12% are non-charged.

Figure 3.31 Peptide generation for the production of antibodies derived from the PnVPP amino acid sequence. The terminal cysteine (grey box) was added to facilitate conjugation to carrier protein. The asterisk (*) indicates the antigenic site as predicted by WebAngis.

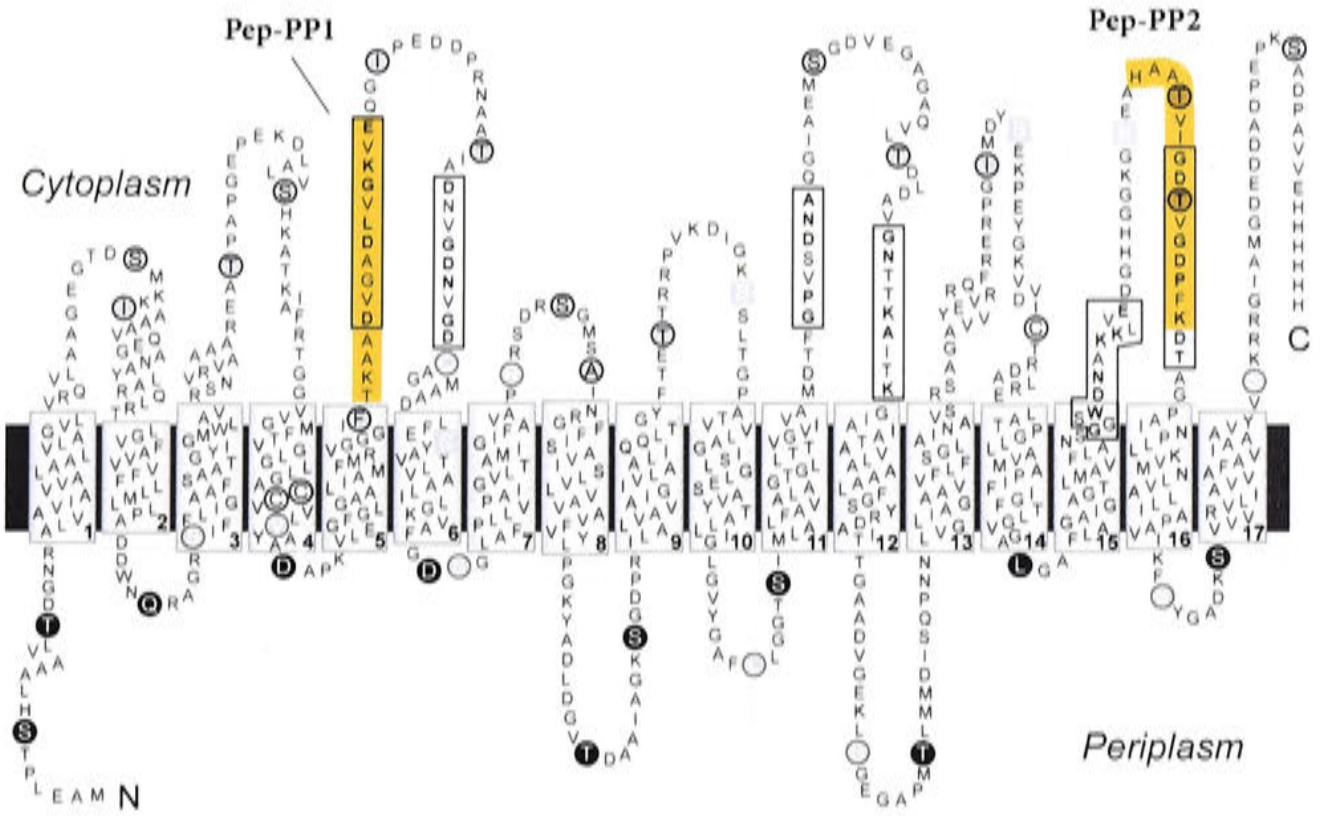


Figure 3.32 Location of peptides within the pyrophosphatase protein. Highlighted regions represent peptide sequences Pep-PP1 and Pep-PP2 that were used for immunisation. (Source:Mimura *et al.*, 2004)

3.2.9 Production and purification of polyclonal antibodies against two PnVPP peptides

For the production of polyclonal antibodies, four rabbits were immunised with peptides Pep-PP1 (#1 and #2) and Pep-PP2 (#3 and #7) following protocols described in Section 2.5.4.1. Pre-immune sera were collected from all rabbits before the first immunisation. Test bleeds were collected three times, 7 days after each booster injection and response to the peptides was monitored by ELISAs. Figure 3.33 shows the progress of immunisation in rabbit #2 over time. The antibody titer for the final sera was calculated by estimation of the inflection point of the post-immune graph and subsequent interpolation by drawing a line down the x-axis as depicted in Figure 3.34. The titer for the immune serum of rabbit #2 was estimated to be 1 in 1200. Only weak responses were observed for rabbit #3. No reaction to peptides was detected in the other two rabbits after several booster injections.

Immune sera from rabbits #2 and #3 were tested in immunofluorescence (3.2.11) and immunoblot assays (3.2.12). As sera showed significant background labelling antibody purification was carried out as shown in Figure 3.35a, following protocols described in Section 2.5.5 to increase specificity. Collected fractions were subjected to Bradford assays and SDS-PAGE and pooled according to protein levels and purity. The final protein concentration of both antibodies was 0.4 mg/mL as determined by Bradford assay. At a

concentration of 7.5 µg/mL both antibodies showed labelling of *P. nicotianae* zoospores in immunofluorescence assays at a brightness comparable to that of the unpurified antibody.

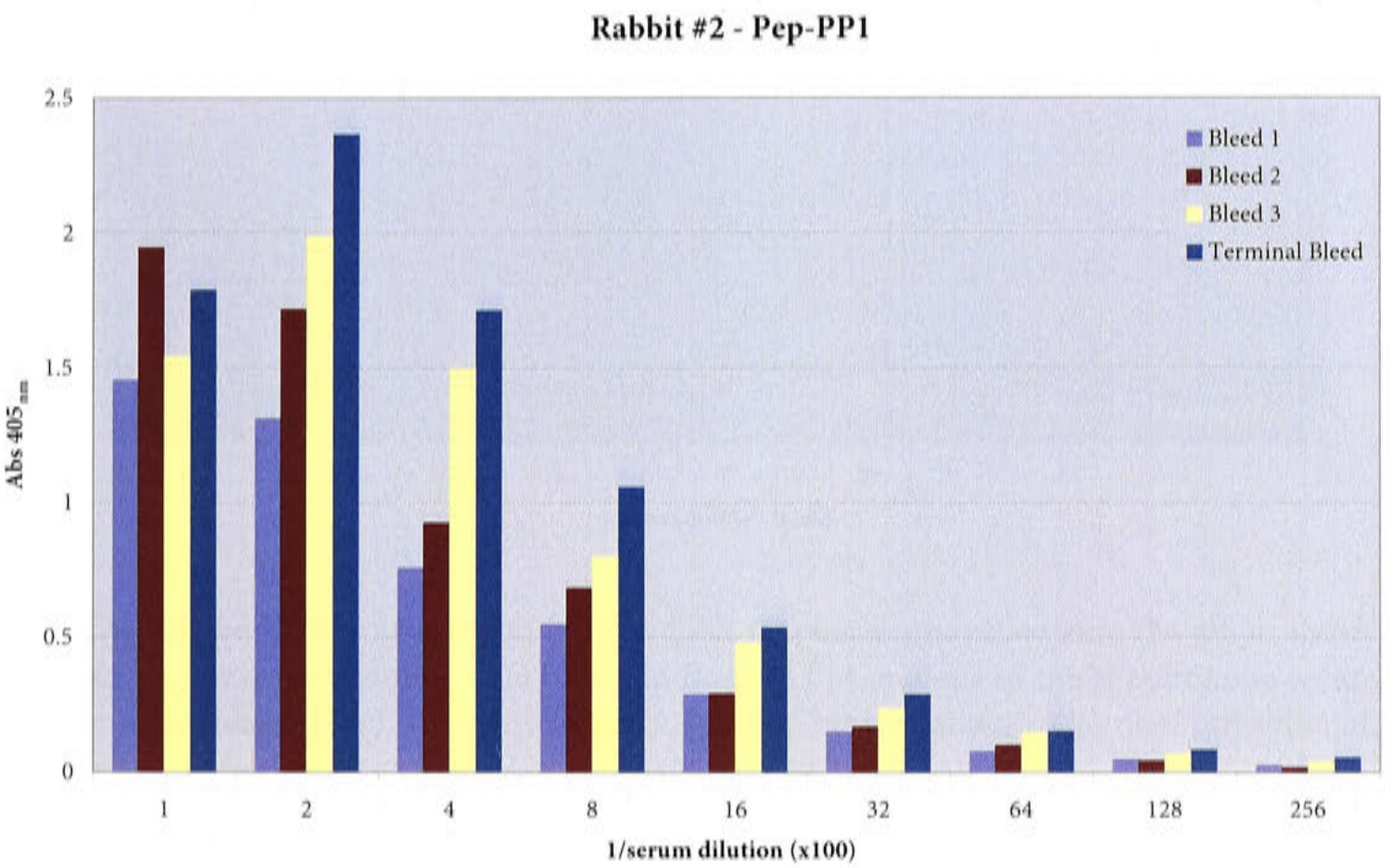


Figure 3.33 Antibody response to immunisation of rabbit #2 with peptide Pep-PP2 as determined by ELISA. The graph shows serial dilutions of three bleeds during the immunisation process as well as the terminal bleed. Absorbance readings shown are after subtraction of pre-immune sera readings. Rabbit #2 reacted well to repeated booster injections with immunising peptide Pep-PP2.

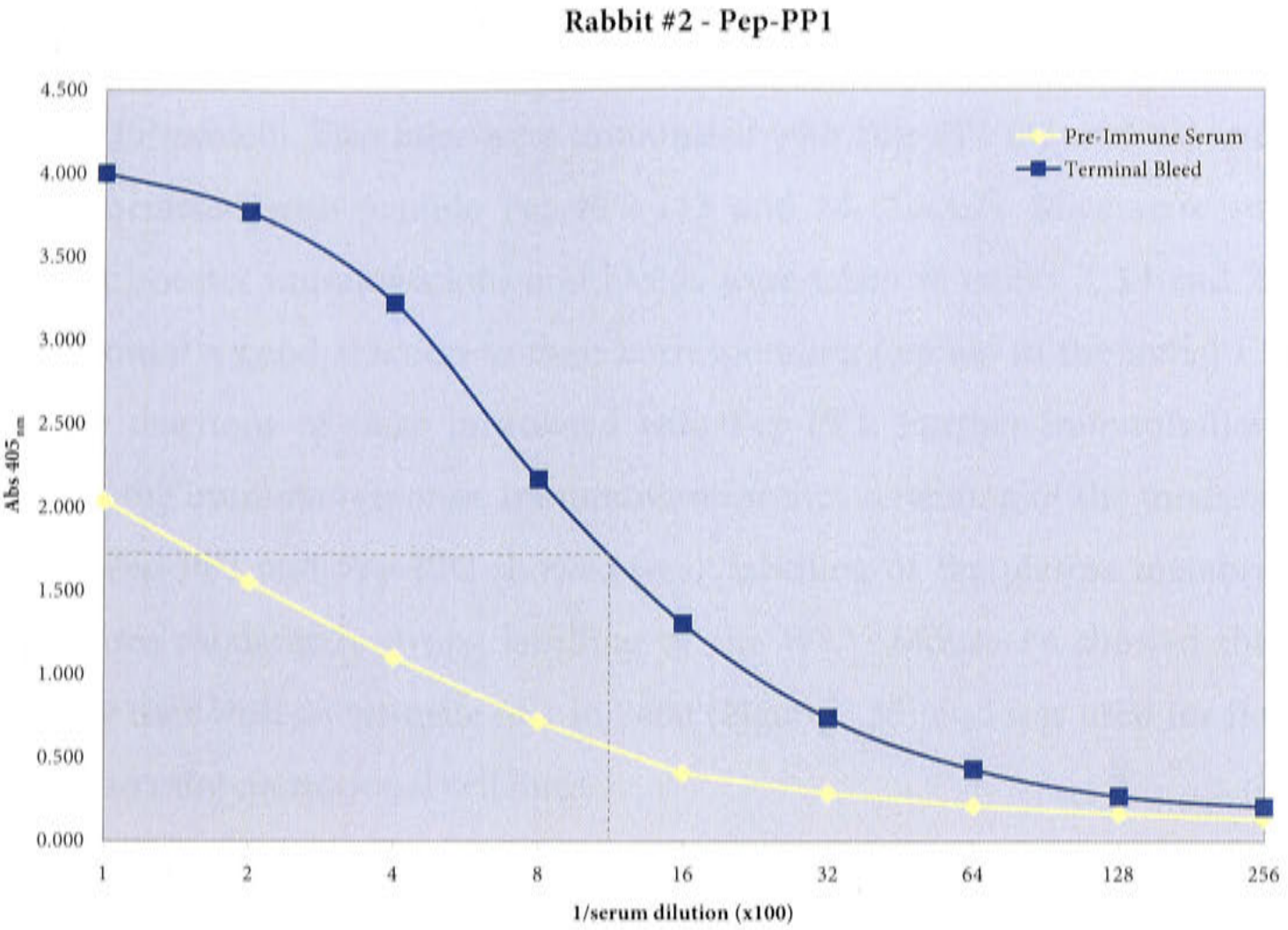


Figure 3.34 Antibody titer estimation of the rabbit #2 post-immune serum. The graph shows a plot of absorbance vs. antiserum dilution data obtained from ELISA analysis of the pre-immune serum (yellow) and the terminal bleed (blue). The grey, dotted line represents the estimated point of inflection of the post-immune graph. Drawing a line to the x-axis interpolates the titer to 1 in 1200.

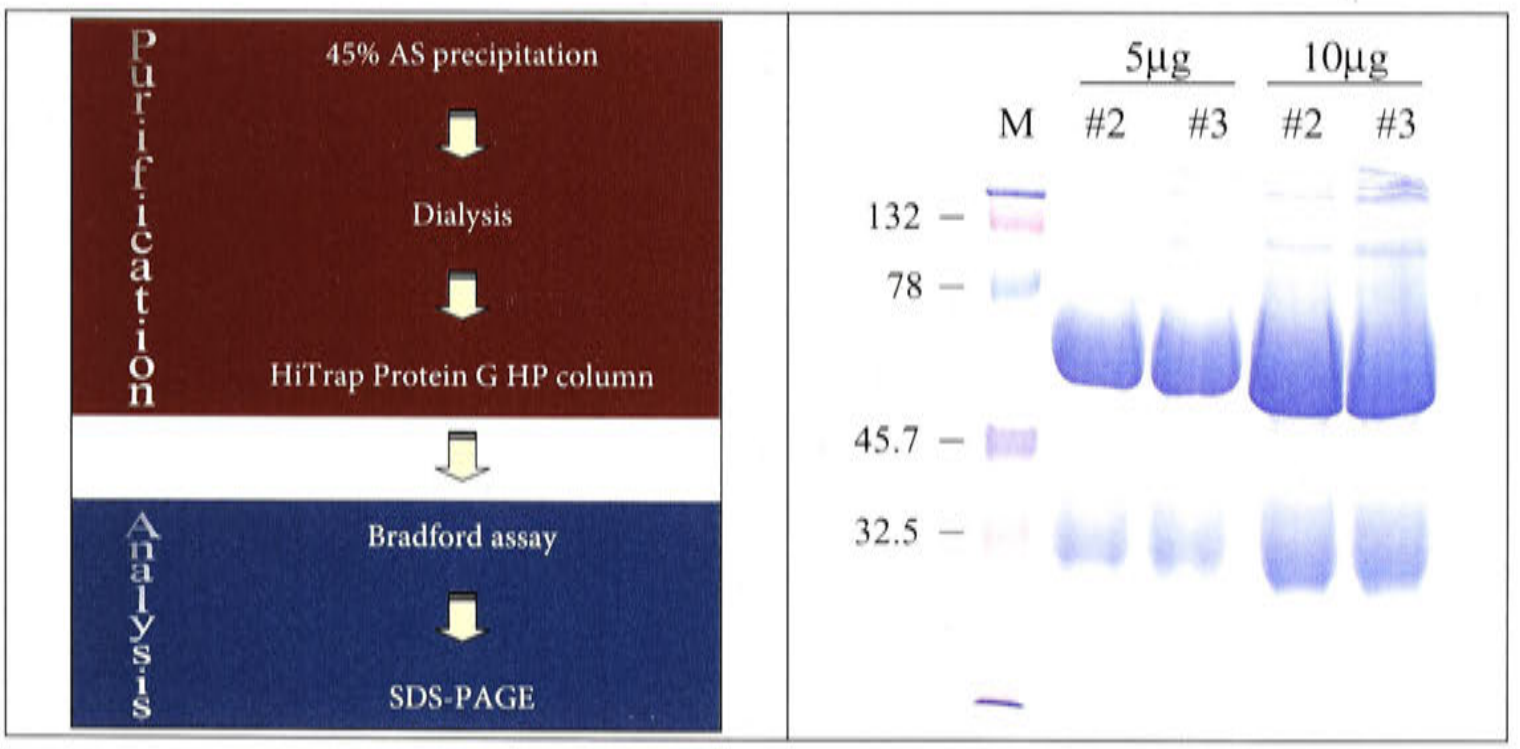


Figure 3.35 Purification of polyclonal antibodies. a) Flow-chart of polyclonal antibody purification, b) SDS-PAGE of purified antibodies after Coomassie Brilliant Blue staining. The final protein concentration after purification was estimated by Bradford analysis and 5 µg and 10 µg protein, respectively, were separated by SDS-PAGE. Numbers on the left indicate position of molecular weight markers in kDa.

3.2.10 Monoclonal antibody production against PnVPP

The aim of the second set of immunisations was to raise monoclonal antibodies against the PnVPP protein. Two mice were immunised with Pep-PP1 (#1 and #2) and two mice were inoculated with peptide Pep-PP2 (#3 and #4; 2.5.4.2). Mice were subjected to frequent booster immunisations and bleeds were taken at weeks 7, 14 and 21. All four mice showed a good reaction to their corresponding peptide in the initial ELISA, with stronger reactions of mice inoculated with Pep-PP2. Further immunisations did not enhance the immune response. Immunofluorescence screening of the mouse sera raised against Pep-PP1 and Pep-PP2 showed weak labelling of the plasma membrane and in some cases moderately strong labelling of the WEV. Mouse #4 showed the strongest antibody titer with an estimate of 1 in 2400 (Figure 3.36) and was used for fusion in the attempt to raise monoclonal cell lines.

Of the 709 cell lines that were screened, 11 were positive after the initial screening in immunofluorescence assays. Those cell lines were subjected to a further round of immunofluorescence screening as well as western immunoblot analysis. Only four clones (2G8, 5D4, 10D5, 1E6) remained positive and those cell lines were cloned and culture supernatant collected for use in further experiments.

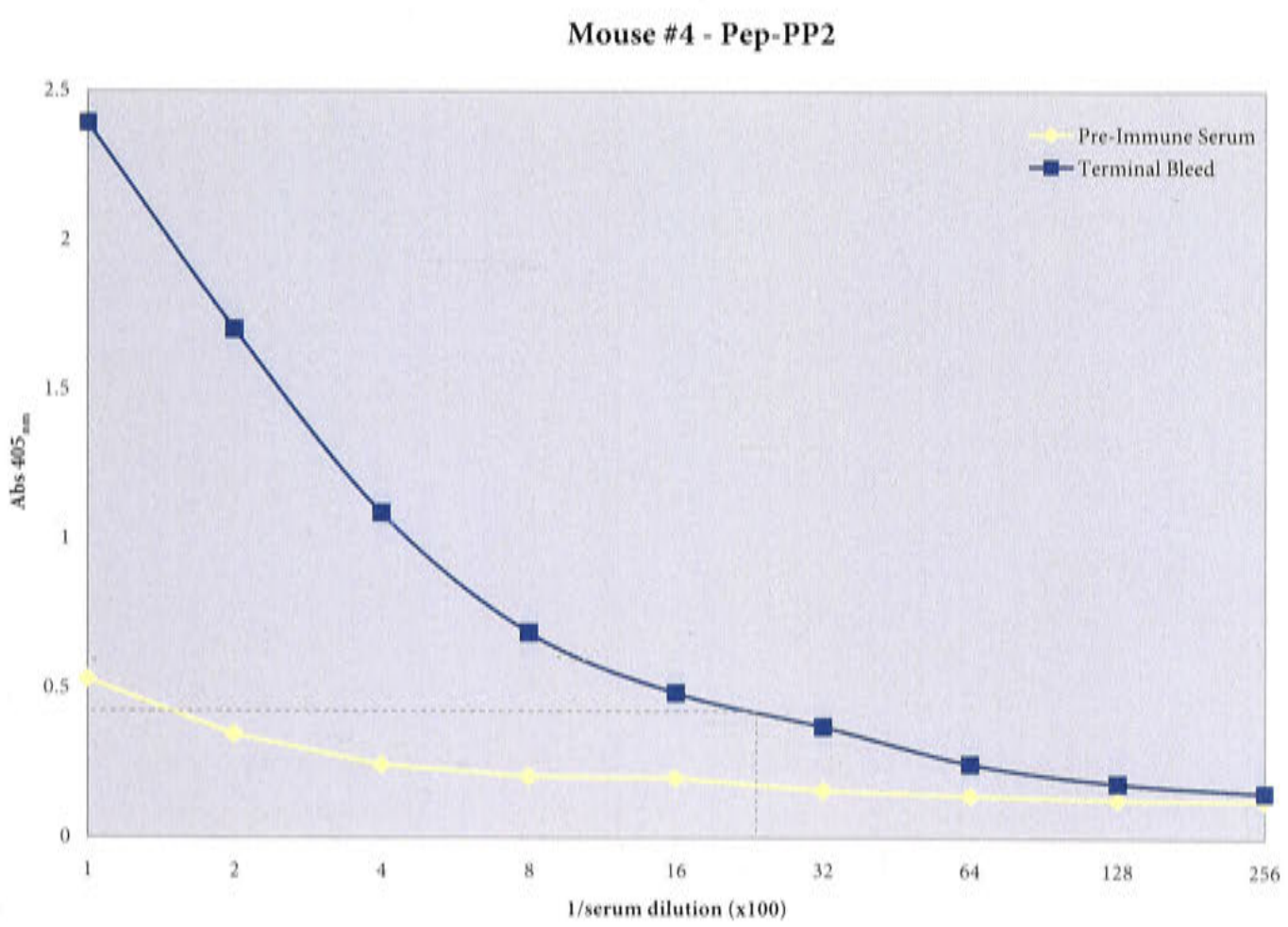


Figure 3.36 Antibody titer of Mouse 4 determined by ELISA. The interpolated titer as determined by the inflection point of the post-immune serum absorbance reading is 1 in 2400.

3.2.11 Immunolocalisation of the V-PPase in *P. nicotianae* zoospores

Immunofluorescence labelling of *P. nicotianae* zoospores with either formaldehyde only or with glutaraldehyde and formaldehyde was carried out. Formaldehyde fixation alone gives the antibodies access to the cytosol and exposes intracellular binding sites due to poor preservation of the plasma membrane. Fixation with a combined solution of formaldehyde and glutaraldehyde preserves an intact plasma membrane and therefore allows surface labelling only (Hardham, 1985).

Sera of rabbits #2 and #3 showed labelling of the plasma membrane as well as the WEV (Figure 3.37a) in zoospores fixed in a combined glutaraldehyde-formaldehyde solution. However, WEV labelling was not always observed in zoospores displaying a WEV in DIC images. No labelling occurred with the pre-immune sera of either rabbit (data not shown). All four monoclonal antibodies showed an identical reaction towards the WEV and the plasma membrane. Figure 3.37b and c show labelling of monoclonal cell lines 5G4 and 1E6.

Labelling of formaldehyde only fixed zoospores showed a variation in labelling pattern. Formaldehyde fixed zoospores tested with antibodies from rabbits #2 and #3 showed weak spotted labelling within the cytosol. When formaldehyde fixed zoospores were tested with the monoclonal antibodies, a bright peripheral labelling pattern was observed (data not shown).

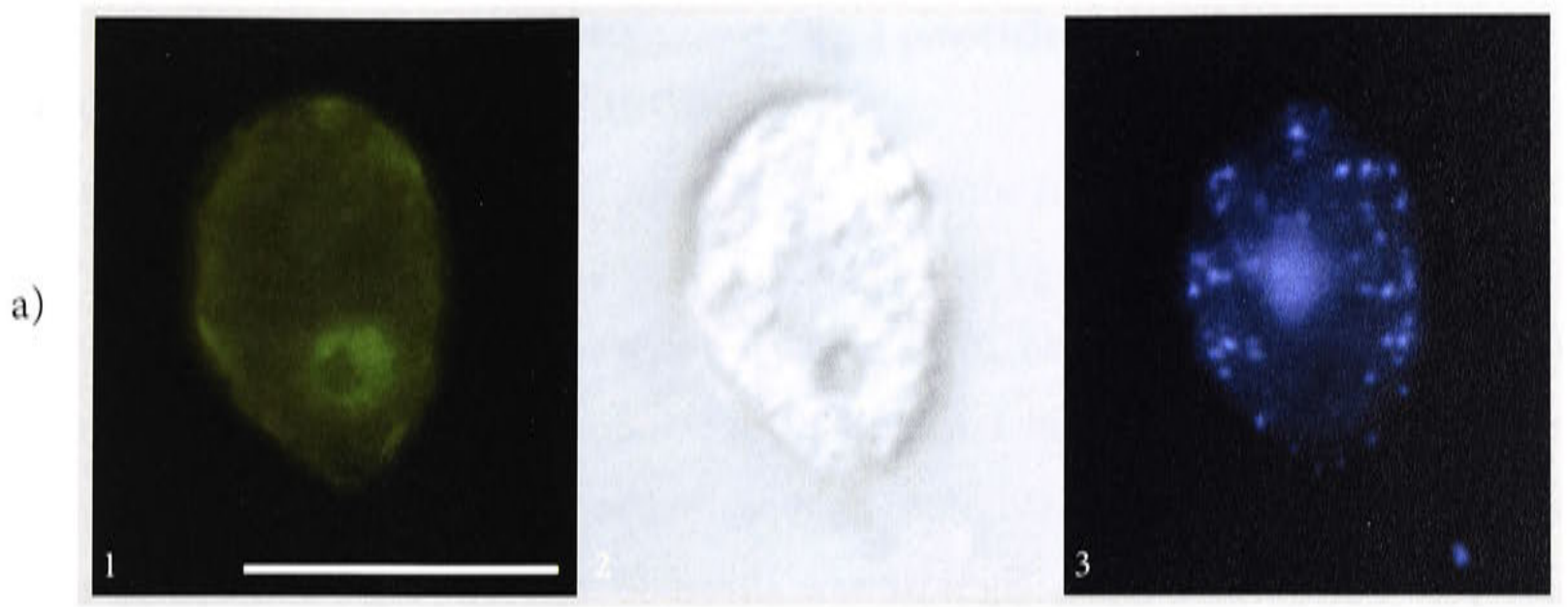


Figure 3.37 Immunolocalisation of PnVPP in zoospores of *P. nicotianae* fixed in glutaraldehyde plus formaldehyde. (a) Zoospores labelled with 10 µg/mL antipeptide-antibodies of Rabbit #2 react with the WEV and the plasma membrane, (b) and (c) labelling of zoospores with monoclonal antibodies 5D4 and 1E6, respectively display the same labelling pattern of the WEV and the plasma membrane. Each set of micrographs show the same cell. 1 - FITC labelling, 2 - DIC, 3 - DAPI staining. Scale bar = 10 µm

3.2.12 Western immunoblot analysis and peptide competition assays of polyclonal and monoclonal antibodies

Immunoblots were carried out on protein extracts from zoospores of *P. nicotianae* that had been solubilised in 2D-extraction buffer, separated by SDS-PAGE and blotted onto nitrocellulose membrane as described in Section 2.5.10. No labelling was detected with pre-immune sera of rabbits #2 and #3. A dilution of 1 in 2000 of Rabbit #2 immune serum weakly labelled various bands in the high molecular weight range as well as two strong bands. The predicted size of bands was strongly dependent on the molecular weight marker used. The Kaleidoscope prestained standard gave estimated molecular weights of 78 kDa and 45 kDa (Figure 3.38), respectively, whereas bands were calculated to run at 65 kDa and 36 kDa, respectively, when the Kaleidoscope precision plus marker was used (Figure 3.40).

A peptide competition assay was carried out in order to determine whether the labelled bands were a result of specific binding of the antibody to the target protein or due to non-specific interactions. A 200-fold excess of the immunising peptide to the antibody did not block labelling of any of the bands suggesting non-specific binding of the antibody. Experiments were repeated with purified antibodies at a concentration of 1 µg/mL and increased peptide excess (up to 500-fold). No decrease in signal intensity was detected (Figure 3.38). Rabbit #3 did not lead to immunolabelling of immunoblots.

Peptide competition dot-blots showed, however, that immune sera of rabbit #2 and rabbit #3 recognise their respective peptides and that labelling can be abolished with a 500-fold excess of the immunising peptide (Figure 3.39). Subsequently, an immunoblot was carried out with immune serum of rabbit #2 on protein of all four life cycle stages of *P. nicotianae*. Labelling was consistent in all developmental stages (i.e. two bands at ~65/75 kDa and ~36/45 kDa), but no decrease in signal intensity was observed after peptide competition using a 500-fold excess of peptide Pep-PP1 (data not shown).

Two of the four monoclonal antibodies (2G8 and 5D4) gave labelling of immunoblots. 2G8 labelled a single band around 48 kDa whereas 5D4 showed labelling of multiple bands in a range of 45 – 200 kDa (Figure 3.40). Peptide competition assay of antibody 2G8 did not show a decrease in labelling intensity. However, it was observed that 2G8 reacted with an additional band at ~75 kDa in this experiment (Figure 3.41).

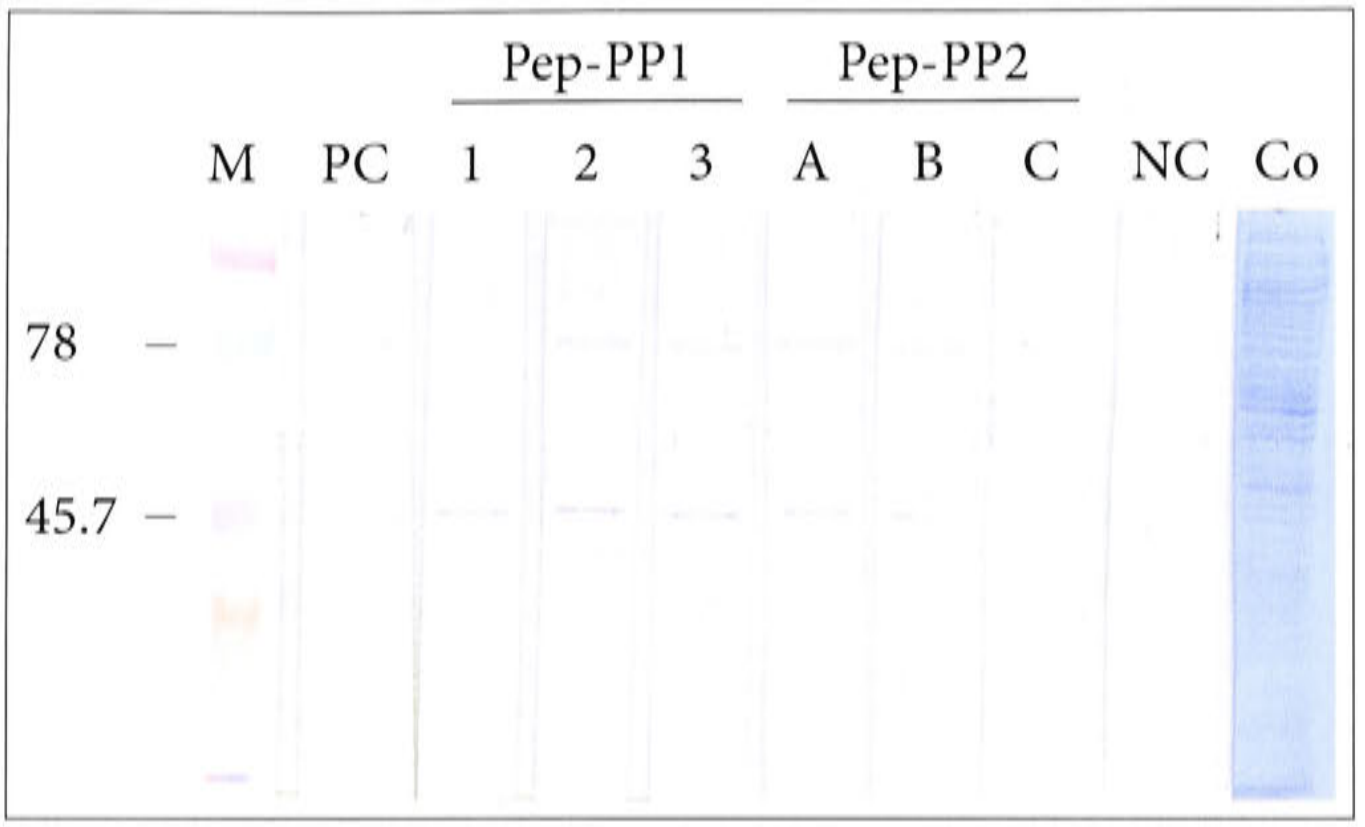


Figure 3.38 Immunoblots of peptide competition assay of *P. nicotianae* zoospore proteins separated by 10% SDS PAGE after incubation with anti-peptide antibodies of rabbit #2 (1 µg/mL). The antibody reacts with two bands ~ 78 kDa and ~45 kDa (lane PC). Pre-incubation of the antibody with peptide Pep-PP1 (lane 1 50-fold excess, lane 2 200-fold excess, lane 3 500-fold excess) does not show a reduction in signal intensity. Control peptide Pep-PP2 (lane A-C, peptide excess as above) does not affect binding efficiency of the antibody. Numbers on the left indicate position of molecular weight markers in kDa. M – Kaleidoscope marker, NC – 2nd antibody only, Co – Coomassie Brilliant Blue staining

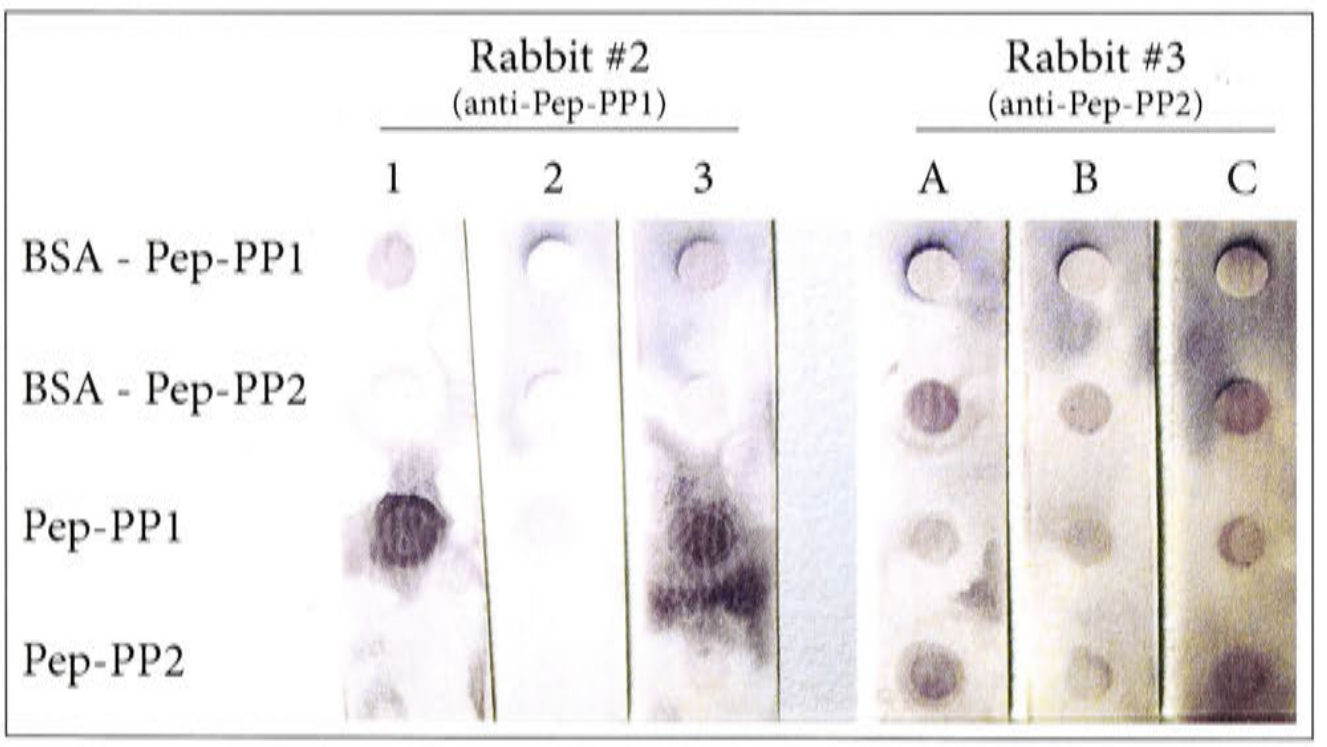


Figure 3.39 Immunodot blot of peptide competition assay of BSA-conjugated and unconjugated peptides labelled with immune sera from rabbits #2 and #3. Both sera show reduced antibody binding when incubated with their respective peptide, whereas control peptides do not affect binding. Columns 1-3 show results after labelling with 1 µg/mL antibody of rabbit #2. Column 1 displays antibody only labelling, whereas Columns 2 and 3 show reactions after pre-incubation with a 500-fold excess of peptides Pep-PP1 and Pep-PP2, respectively. Antibody binding is abolished by incubation with peptide Pep-PP1 (row Pep-PP1 / column 2). Columns A-C display antibody labelling of rabbit #3 (5 µg/mL). In column A the antibody reacts with its peptide in BSA-conjugated and unconjugated form. Reaction is diminished by pre-incubation with peptide Pep-PP2 (Column B) but not by peptide Pep-PP1 (Column C).

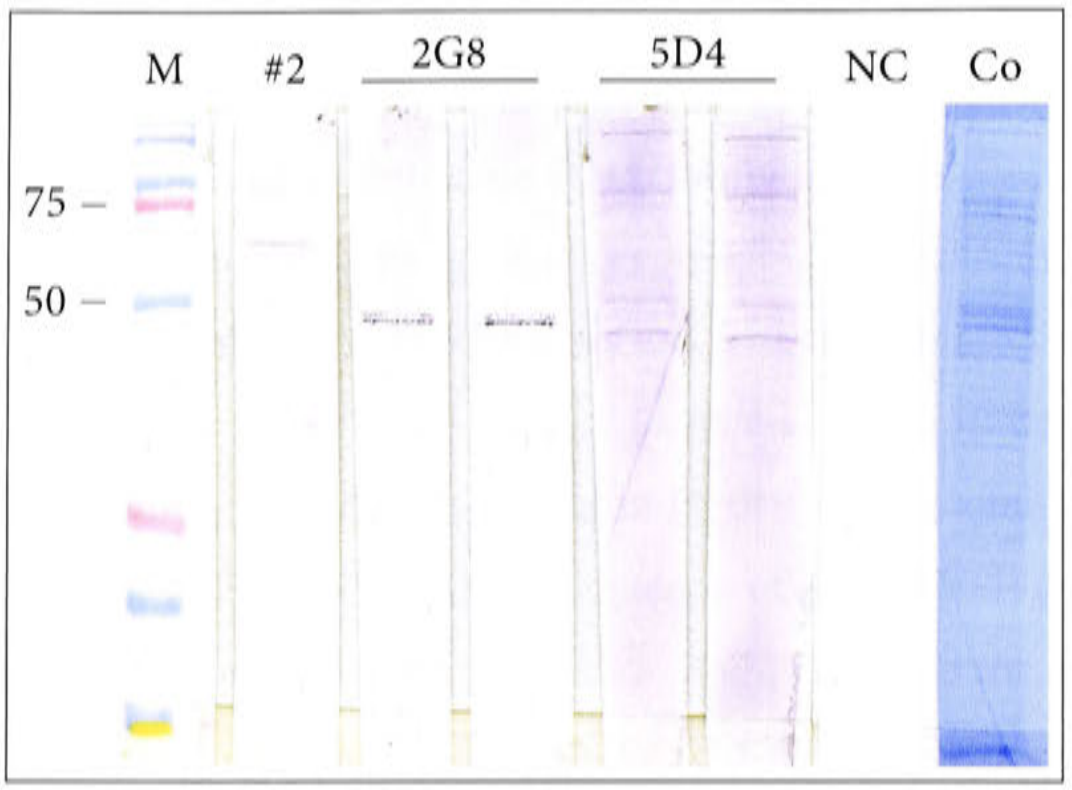


Figure 3.40 Immunoblots of *P. nicotianae* zoospore protein extract after incubation with anti-peptide antibodies of mouse #4 (2G8 and 5D4). Monoclonal cell line 2G8 reacts with a ~48 kDa band, whereas 5D4 labels a range of bands. Immune serum of Rabbit #2 (lane #2) was run for comparison of labelling patterns. Numbers on the left indicate position of molecular weight markers in kDa. M – Kaleidoscope marker, NC – 2nd antibody only, Co – Coomassie staining

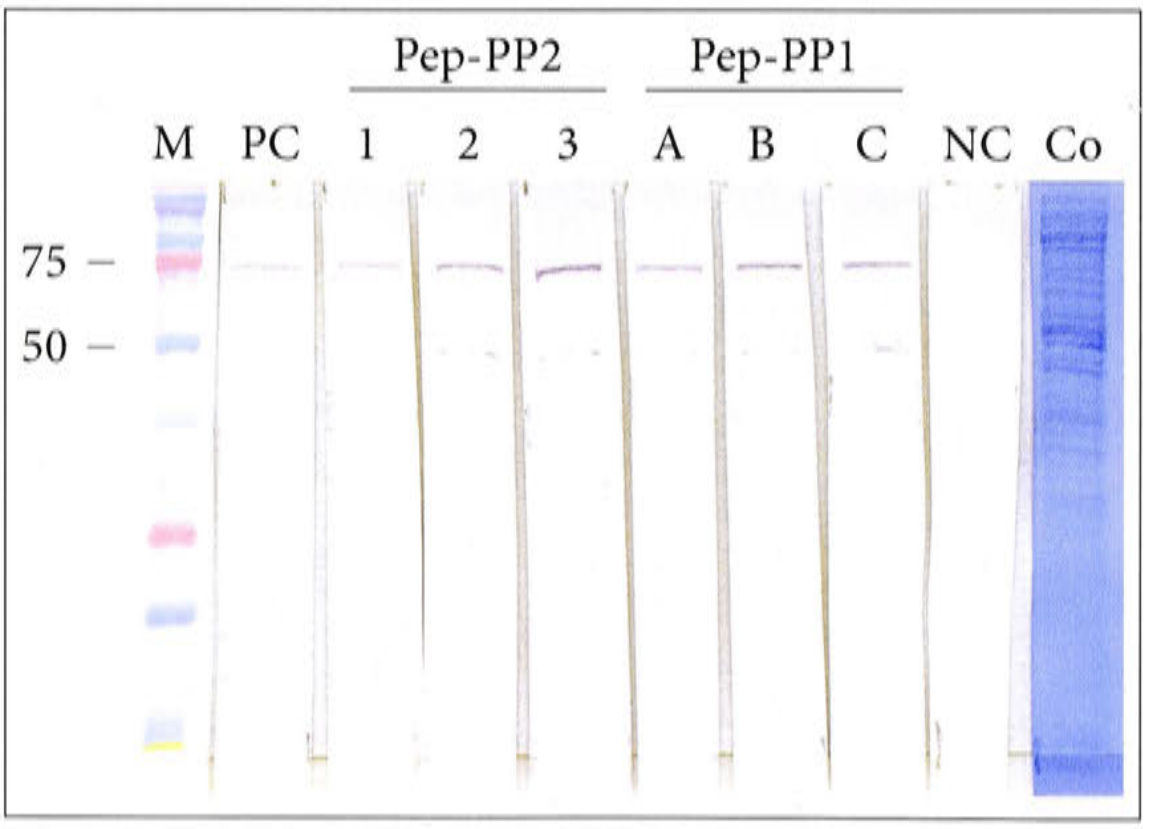


Figure 3.41 Immunoblots of peptide competition assay of *P. nicotianae* zoospore protein extract separated by 10% SDS PAGE after incubation with monoclonal antibody 2G8. The antibody reacts with two bands ~ 75 kDa and ~48 kDa (lane PC). Pre-incubation of the antibody with peptide Pep-PP2 (lane 1 50-fold excess, lane 2 200-fold excess, lane 3 500-fold excess) does not show a reduction in signal intensity. Control peptide Pep-PP1 (lane A-C, peptide excess as above) does not affect binding efficiency of the antibody. Numbers on the left indicate position of molecular weight markers in kDa. M – Kaleidoscope marker, NC – 2nd antibody only, Co – Coomassie staining

3.2.13 Construction of a dsRNA construct for transformation assays in *P. nicotianae*

RNA interference (RNAi) is commonly used to analyse gene function by introduction of dsRNA constructs which lead to antisense RNA inhibition and a subsequent knock-down of the targeted gene and its protein expression levels. A dsRNA construct expressing a 303 bp fragment of the *PnVPP* gene was first cloned into vector pTEP5 and then transferred into vector pTH210 for use in *P. nicotianae* transformation experiments. A flow chart of the cloning procedure is shown in Figure 3.42. To obtain a plasmid that expresses dsRNA duplexes, a partial cDNA sequence of the *PnVPP* gene as well as a modified Kan^R gene was cloned into vector pTEP5. Both, Kan^R and the dsRNA sequence were PCR amplified using primers KANF/KANR and T3/PP_dsRNA rev, respectively. A triple ligation was performed using the linearised vector pTEP5 (*Spe*I), the *Cla*I digested Kan^R PCR product and the double-digested (*Spe*I/*Cla*I) *PnVPP* dsRNA insert. A schematic representation of the final *PnVPP* dsRNA construct is presented in Figure 3.43. Successful cloning and correct insertion of the inserts was confirmed by colony PCR, control digests of the isolated plasmid DNA, subsequent sequencing and alignment of the obtained sequence with the *PnVPP* gene (Figure 3.44).

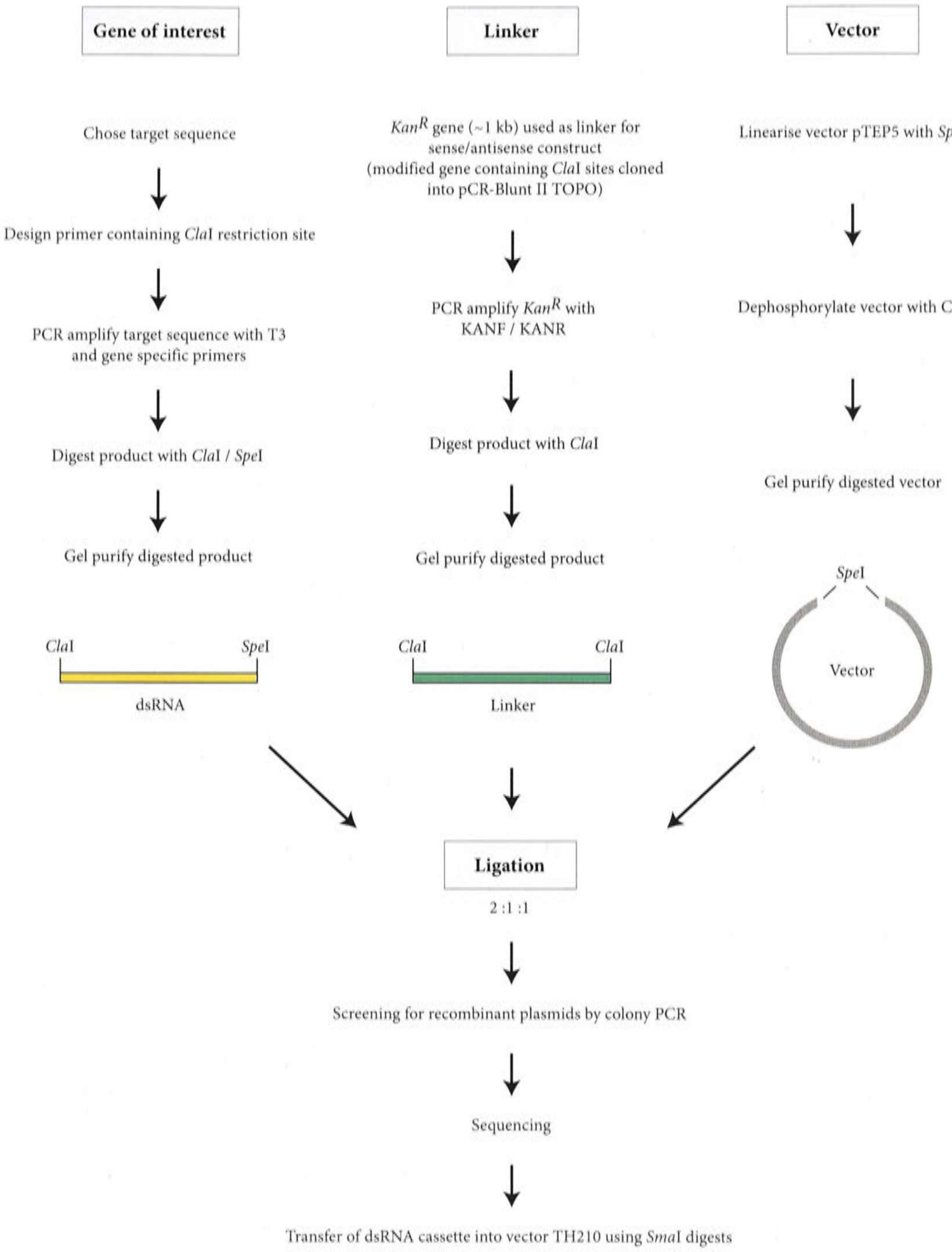


Figure 3.42 Flow-chart for the procedure of cloning a dsRNA construct for the use in *P. nicotianae* transformation experiments. For further details, see text.

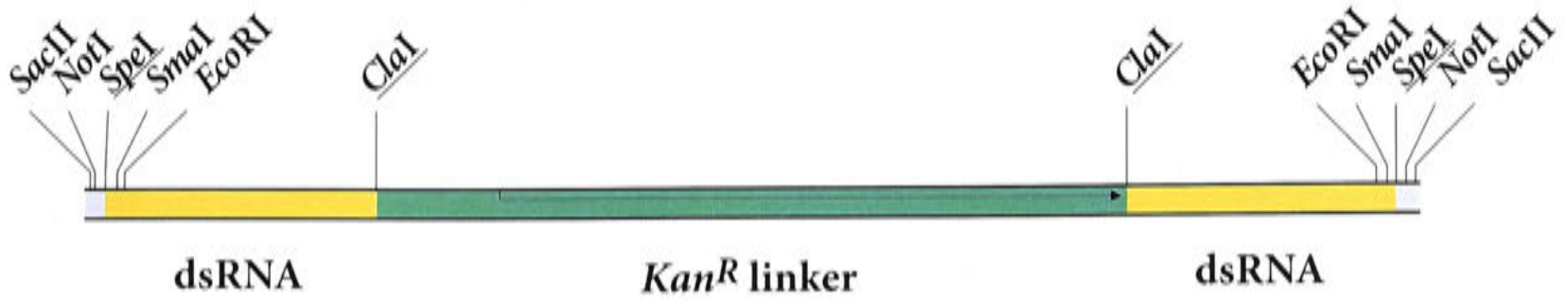


Figure 3.43 Schematic representation of the *PnVPP* dsRNA plasmid. The yellow regions denote the dsRNA sequence, green shows the *Kan*^{R} linker gene and grey the vector. Restriction enzyme sites utilised for cloning into vector pTEP5 are underlined.

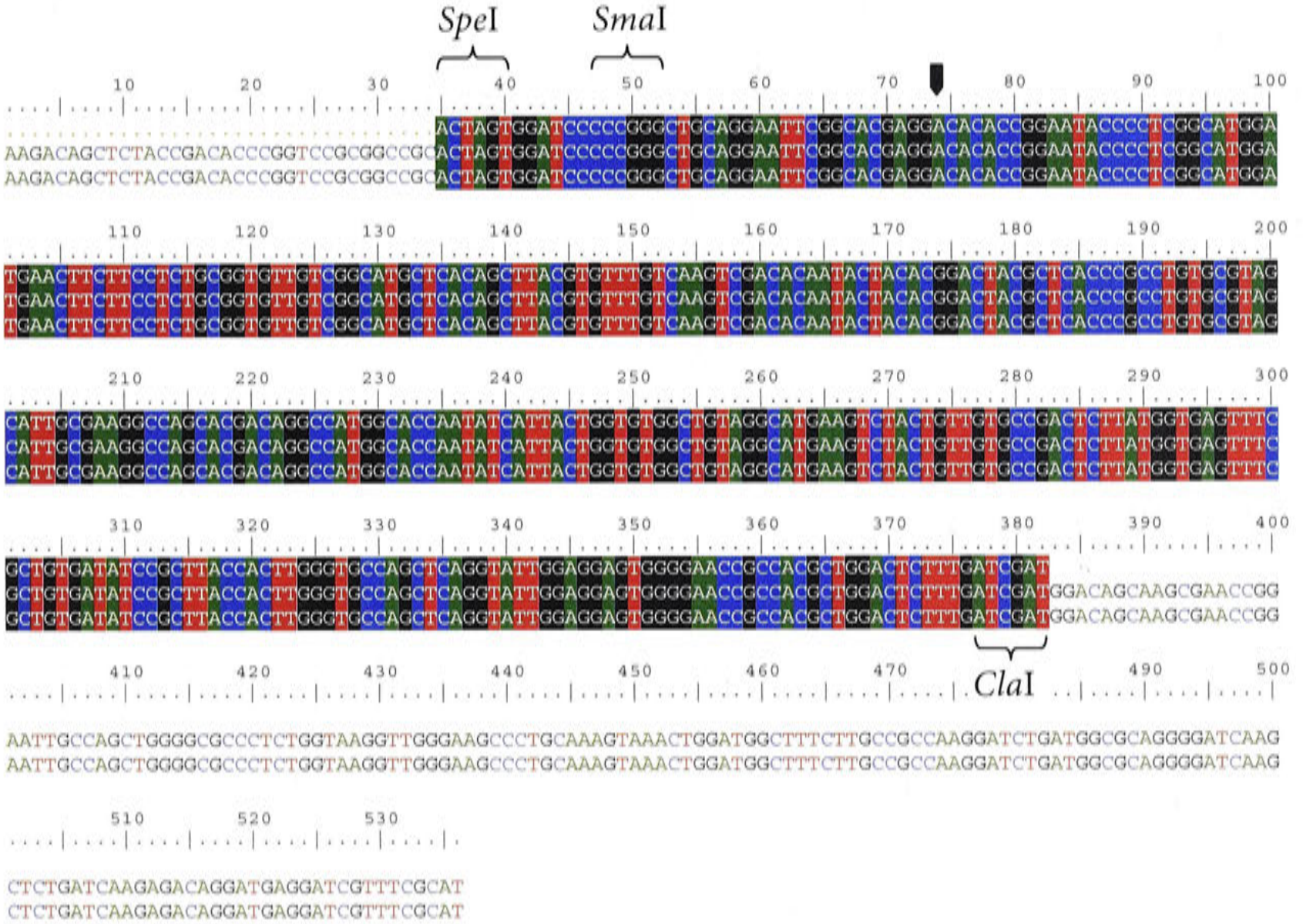


Figure 3.44 Alignment of the selected *PnVPP* dsRNA sequence with sequencing results of cloned plasmid. The dsRNA sequence (upper row) shows relevant 5' and 3' restriction enzyme sites obtained by using primer T3 (5') and a specifically designed GSP incorporating the *Cla*I site (3') for PCR amplification. The black arrow indicates the start of the actual *PnVPP* sequence. The alignment shows no mismatches between the predicted sequence and the sequencing data obtained from the plasmid.

3.3 Discussion

V-PPase activity was originally reported in *R. rubrum* chromatophores in 1966 (Baltscheffsky *et al.*, 1966), however, a gene encoding a V-PPase was not cloned until 1992 (Sarafian *et al.*, 1992). Since then, V-PPases have been studied extensively and have been identified in plant vacuoles, acidocalcisomes (acidic calcium containing organelles) of protozoa, bacteria, slime molds and algae. Recently, a Type II V-PPase was identified in animal cells.

This chapter described the cloning and characterisation of the gene encoding the V-PPase protein in *P. nicotianae* and the localisation of the V-PPase in zoospores. The study is the first report of a cloned and characterised V-PPase gene in a *Phytophthora* species.

3.3.1 Cloning, structural organisation and expression of the *PnVPP* gene in *P. nicotianae*

Southern blot analysis showed that the *P. nicotianae* genome contains a single copy of the V-PPase gene. RNA blot analysis of *PnVPP* revealed an ~ 2400 bp transcript which is present in vegetative hyphae, zoospores and germinated cyst, but is not detected in sporulating hyphae. The transcript is up-regulated in zoospores and germinated cysts.

Initial comparison of the *PnVPP* genomic clone and the DS311 cDNA sequence showed the presence of a single 70 bp intron and two alternative in-frame translation start sites positioned 285 bp apart, corresponding to a difference of 95 aa or 10 kDa in the PnVPP protein. As V-PPase proteins from other organisms range from 686 – 1044 aa in size (Table 3.1), typical polypeptide length did not provide a clue as to which ATG corresponded to the coding region. Re-screening the λ-ZAP *P. nicotianae* zoospore cDNA library identified two cDNA clones, A-5 and E-20, which included approximately 750 bp upstream of DS311, but which still did not cover the complete *PnVPP* transcript. In general, a good representational primary library size consists of at least ~1 × 10^6 clones (Short *et al.*, 1988). However, the construction of high-quality cDNA libraries is highly dependent on the quality of isolated RNA and the efficiency of cDNA synthesis, and libraries often have a high content of non-full length cDNA clones (Lu *et al.*, 1997). The *PnVPP* transcript is fairly large with ~2400 bp and as re-screening of approximately 2.5 x 10^5 clones revealed only one truncated clone, it was concluded that the library most likely does not contain any full-length cDNA clones of this gene. Therefore, a 5' RACE

approach was employed to amplify the unknown 5' region of the *PnVPP* gene. As the *PnVPP* transcript is up-regulated in zoospores, this developmental stage was chosen for RNA isolation to enhance the efficiency of amplification. Experiments were repeated three times and all PCR amplifications resulted in a PCR product of ~500 bp. Sequencing of these products resolved the question as to which ATG was the correct translational start site and additionally gave some information on the 5' UTR, the beginning of which might reflect the transcription start site (TSS) of the gene. However, in order to confirm the TSS other experiments, for example, primer extension or nuclease protection assays, need to be carried out. Due to the uncertainty of the actual TSS, positions mentioned in the text below will refer to the translational start codon and not to the TSS in the *PnVPP* gene. It was inferred from the sequence obtained from the 5' RACE product that the second ATG (at position 729 in the genomic clone in Figure 3.22) serves as the start site for the expression of this gene, yielding a 2403 bp transcript which encodes an 801 aa (84 kDa) polypeptide.

The likelihood that the second ATG was the translation start site was strengthened by the identification of motifs similar to elements typical of eukaryotic promoters in the sequence upstream of the second ATG. The putative 5' UTR contained sequences similar to a TATA-box, a CAAT motif and CT-rich regions. The TATA-box is commonly located within 35 bp of the TSS (Lodish, 2000) and was found at position -237 upstream of the ATG in the 5' UTR of the *PnVPP* gene. Other promoter-proximal elements that regulate gene expression, such as CAAT motifs, GC box and CT-rich regions, are generally located within 200 bp of the TSS (Lodish, 2000). In the 5' flanking region of *PnVPP* two CAAT-like motifs (-280 bp and -140 bp from the ATG) and two CT-rich elements (-73 to -54 bp and -44 to -23 bp from the ATG) are present. A conserved 16 bp motif with the core sequence GCYCATTYYNCAWTTTNYY situated within the first 100 bp upstream of the translational start codon of many oomycete genes (Pieterse *et al.*, 1994; McLeod *et al.*, 2004) was found at position -79 of the *PnVPP* gene.

Sequences resembling AT-rich polyadenylation signals (consensus AATAAA) are commonly found downstream of oomycete ORFs (Pieterse *et al.*, 1994; Kamoun, 2003; Win *et al.*, 2006). A putative polyadenylation signal AAACAA was found 36 bp upstream from the poly(A) tail.

Approximately 68% of filamentous fungal genes contain introns (Pieterse *et al.*, 1994), but in contrast, the majority of oomycete genes examined to date do not contain introns. The analysis of 63 *Phytophthora* genes in 2002 revealed that just 33% (21 genes) contained a total of 32 introns, or an average of 1.5 introns per gene (Kamoun, 2003). A more recent study of *P. infestans* genes showed that among the 33 genes for which the full genome sequence was recovered, eight genes (27%) had introns (six genes had one intron each and two genes had two introns each). Furthermore, a total of nine introns was found in seven of the 68 partially covered genes (Win *et al.*, 2006). The average length of introns is noted with approximately 80 bp (Kamoun, 2003; Win *et al.*, 2006). The G+C content of introns was reported as 44% compared to a G+C content in coding regions of 57% (Win *et al.*, 2006). This conforms with the characteristics of the *PnVPP* gene, which shows a G+C content of 40% in the intron and 56% in the coding region, respectively.

Only seven of the 23 cloned and characterised V-PPase genes (Table 3.1) have both cDNA and genome sequence available. Interestingly, only one V-PPase gene (*T. gondii*) other than the *PnVPP* gene described here contains one or more introns (Drozdowicz *et al.*, 2003).

The single intron in the *PnVPP* gene showed splice junctions of GT/GGTAC (5') and TA/GAT (3'), respectively. This does not conform with the consensus from 55 other *P. infestans* introns, which were recorded as $A_{48}G_{64}/G_{100}T_{100}A_{66}A_{44}G_{86}T_{60}$ for the donor sites and $C_{58}A_{100}G_{100}/G_{46}$ for the acceptor site (Cvitanich and Judelson, 2003a; Kamoun, 2003). However, it was reported that the intron splice junctions of *P. infestans* did not show similarity to those of *P. sojae* and *P. ramorum* introns, suggesting that intron splice junction sequences are diverse within *Phytophthora* species (Win *et al.*, 2006). Another motif, CTAAC, which is present in the intron of the *PnVPP* gene, is believed to be important for lariat formation during splicing has been reported in several intron sequences (Kamoun, 2003).

3.3.2 Characteristics of the PnVPP protein – conserved domains and its structure

The *PnVPP* gene encodes a V-PPase protein with homology to AVP2 from *A. thaliana*. A CDD search (Marchler-Bauer *et al.*, 2005) with the translated sequence from the full length genomic clone of *PnVPP*, using the BLASTP alignment, produced hits containing

conserved domains within the inorganic H^+-pyrophosphatase family. This family includes members from plants, archaea and parasitic protists; some representatives of this group are: AVP1 and AVP2 from *A. thaliana* (Sarafian *et al.*, 1992; Drozdowicz *et al.*, 2000), PfVP1 and PfVP2 from *P. falciparum* (McIntosh *et al.*, 2001), CPP1 from *C. corallina* (Nakanishi *et al.*, 1999), PVP from *P. aerophilum* (Drozdowicz *et al.*, 1999) and RVP *R. rubrum* (Baltscheffsky *et al.*, 1998).

V-PPases across all organism analysed to date exhibit striking sequence conservation in a number of primarily cytosolic loops. The putative catalytic motif Dx_7KxE, which is similar to the Ex_7KxE motif found in all sPPases and membrane associated H^+-PPases (Cooperman *et al.*, 1992), is located in loop *f* in the *P. nicotianae* protein (see Figure 3.24 for references to TMD and loops). The three residues D, K and E in this motif have been shown to be essential for catalysis as judged by the loss of PP_i hydrolytic activity and H^+ translocation upon non-conservative substitution by site-directed mutagenesis (Nakanishi *et al.*, 2001; McIntosh and Vaidya, 2002). Furthermore, substitutions of those residues rendered the protein incapable of H^+-translocation, but retained 25-50% of hydrolytic activity, indicating that these residues are not only involved in substrate binding but also in the coupling of PP_i hydrolysis to H^+-translocation (Nakanishi *et al.*, 2001; McIntosh and Vaidya, 2002). The two transmembrane proximal acidic motifs Dx_3Dx_3D were also entirely conserved in PnVPP. The first motif is found adjacent to the catalytic motif Dx_7KxE (amino acid residues 299-307, loop *f*) and the second Dx_3Dx_3D motif is located near the N-terminus (amino acid residues 765-773, loop *p*). Involvement of the acidic residues of this motif in enzyme catalysis, and/or substrate binding was demonstrated by site-directed mutagenesis of the *V. radiata* V-PPase protein. (Nakanishi *et al.*, 2001; McIntosh and Vaidya, 2002).

V-PPases are furthermore divided into Type I and Type II proteins according to their requirement for K^+ (Drozdowicz and Rea, 2001). The BLAST result with the full length protein sequence (801 aa) showed highest homology to *A. thaliana* AVP2, indicating that PnVPP is a Type II V-PPase. Moreover, all of the sequence motifs demonstrated to be necessary for catalysis by Type II V-PPases were conserved in the PnVPP protein. The span-loop interface motif of cytosolic loop *i*, which in all previously characterised Type I V-PPases is T[DE]YYTS, is replaced with T[KQR]YYTD in Type II proteins, and it was inferred that this motif is diagnostic of Type I and Type II V-PPases, respectively

(Drozdowicz and Rea, 2001). In PnVPP, the motif TQYYTD was found in loop *j* (amino acid residues 439-444) in proximity to TMD10, supporting the notion that PnVPP is a Type II protein. A third useful Type I/Type II distinguishing criterion is the conserved motif GNxx[A/K]Ax[T/A/G] located at the interface between the cytosolic loop *i* and its adjoining TMD (Belogurov and Lahti, 2002). It has been shown, by comparison of the amino acid conservation patterns in K^+-dependent and K^+-independent V-PPases, that only two sequence positions, located two amino acids apart, are present where a residue conserved in one type is absent from the equivalent position in all the sequences from the other type. The Ala (A)/Lys (K) position is occupied by Ala in K^+-dependent V-PPases and Lys in K^+-independent V-PPase. Similarly, the Gly(Ala)/Thr position is occupied by Gly or Ala in K^+-dependent V-PPase and Thr in K^+-independent V-PPase (Belogurov and Lahti, 2002). Moreover, it was demonstrated that substitution of alanine by lysine in the Ala/Lys position of a Type I V-PPase abolishes the K^+-dependence of the enzyme, thus indicating that this position is sufficient for the classification of V-PPases (Belogurov and Lahti, 2002). In PnVPP a sequence corresponding to the proposed conserved motif was found at position 563 to 570 in the large cytosolic loop *j* between TMD10 and TMD11, containing amino acid residues lysine and threonine in the positions of interest. Taken together, it can be inferred from the structural features of the *P. nicotianae* V-PPase protein that the enzyme falls into the class of Type II, K^+-independent H^+-translocating pyrophosphatases.

To date, there has been no report on the three-dimensional (3D)-structure of a V-PPase protein. Obtaining diffraction-quality single crystals of membrane proteins in general has been proven to be extremely difficult. Relative to soluble proteins (~30,000 structures in the Protein Data Bank [PDB]), only approximately 60 unique structures from transmembrane proteins are available (Guan *et al.*, 2006). The reason for this is the limited ability to manipulate proteins bearing hydrophobic/amphiphilic surfaces that are usually enveloped with membrane lipid (Caffrey, 2003).

Due to these shortcomings, recent studies focussed on the evaluation of potentially functional residues, secondary structure and the prediction of the tertiary structure. Predictions on the secondary membrane topology and the functional unit of the V-PPase protein have been made by site-direct mutagenesis, cross-linking studies, gel-filtration chromatography, radiation inactivation and thermo-inactivation (reviewed in Mimura *et*

al., 2005), and furthermore by hydropathy analysis, computer program predictions and cysteine-scanning mutagenesis (Mimura *et al.*, 2004). A number of TMD prediction programs were used in this study in combination with a hydropathy plot. Topology models show a highly hydrophobic enzyme with almost half of its residues predicted to be situated within the membrane. The majority of topology prediction programs used predicted a highly similar overall topology with a very comparable residue composition of the core regions. Two exceptions however are TMD7 and TMD17: TMD7 was predicted by only 5 out of 8 programs (62.5%), and TMD17 by 6 out of 8 programs (75%). Based on a recent study on *S. coelicolor*, it was concluded that PnVPP most likely contains all 17 TMDs, with the N-terminus facing the cytosol and the C-terminus facing the vacuolar lumen. In this study, the topology of the 794 aa V-PPase protein of *S. coelicolor* was determined by a combination of cysteine-scanning mutagenesis and the use of sulphhydryl reagents (Mimura *et al.*, 2004). It was revealed that *S. coelicolor* V-PPase contains 17 TMDs and that several conserved segments, such as the substrate-binding domains, are exposed to the cytoplasm. As the *S. coelicolor* protein features a long additional sequence that includes a TMD at the C-terminus, it was proposed that the basic structure of V-PPases which lack the extra sequence has 16 TMDs with the C-terminus exposed to the vacuolar lumen (Mimura *et al.*, 2004). This, however, is in contrast to other reports which favour a 15 TMD model and drop the predicted 16th TMD, based on biochemical evidence which supports a cytosolic location of the C-terminal region (Knight *et al.*, 1996; Zhen *et al.*, 1997; Schultz and Baltscheffsky, 2003).

Comparison of the amino acid sequences of PnVPP (801 aa) and ScPP shows indeed a much longer C-terminus for the *S. coelicolor* protein, however, the *P. nicotianae* sequence features a longer N-terminus instead. The putative catalytic motifs Dx_7KxE and Dx_3Dx_3D are located in cytosolic loop *e* (between TMD5 and TMD6) in *S. coelicolor* but are found in cytosolic loop *f* (between TMD6 and TMD7) in *P. nicotianae*, indicating that the *P. nicotianae* protein has an additional TMD in the N-terminal region and therefore shifting the terminus from the cytosol to the vacuolar lumen. TMD7 was confirmed as an authentic domain by comparison of the *P. nicotianae* amino acid sequence with several structure models. It has been shown that the acidic motif Dx_3Dx_3D is located on the TMD/loop interface (Schultz and Baltscheffsky, 2003; Mimura *et al.*, 2004), thus supporting the presence of TMD7 in the PnVPP protein as TMD8 is too distant with approximately 40 aa downstream of this motif. Furthermore, this topology is also in

agreement with the required cytosolic orientation of certain downstream loops, thought to be involved in catalytic activity and H^+-translocation. The existence of TMD17 remains controversial. Comparison with the ScPP model and the proximity of the second Dx_3Dx_3D motif to the TMD favours its existence. However, this would locate the short C-terminal tail of the protein to the vacuolar lumen and as mentioned above, not all studies agree on the view of lumenal exposure of the C-terminal region. Results of analysis with topology prediction programs remain highly speculative and even though they are useful tools for the prediction of tertiary structures especially in combination with experimental approaches, they cannot give a definite answer. The correct confirmation of the protein, the number of TMDs and the orientation of N- and C-terminal regions can only be obtained by high resolution crystallographic studies.

3.3.3 Immunolocalisation of PnVPP

The aim was to generate polyclonal and monoclonal antibodies against the *P. nicotianae* PnVPP protein to elucidate the location of the enzyme in *P. nicotianae* zoospores. It was attempted to overexpress two segments of the protein in two different expression systems, however, neither approach was successful. The first plasmid contained a 153 bp insert encoding a 51 aa segment of the PnVPP protein. The expressed region encompasses loop *p* which contains the second acidic Dx_3Dx_3D motif and does not contain any TMD. The second plasmid was larger with a 453 bp insert that corresponded to 152 aa, encompassing loops *l-n* and TMDs13 and 14. Expression of the proteins was not observed even after thorough experimental evaluation, considering influencing factors such as IPTG concentration for induction, host strain, temperature and induction time. Overexpression of membrane proteins has always been considered rather challenging. Membrane proteins often induce toxic effects which can lead to the death of the microorganism, most likely due to the association with or incorporation into vital membrane systems (Montigny *et al.*, 2004). This explanation, however, seems unlikely in this study, as the growth of cultures was not impaired at any time. Also, the first small segment does not contain any hydrophobic transmembrane segments. It has also been shown that the choice of vector/overexpression system and host strain is crucial in overexpression of transmembrane proteins. One study, for example, demonstrated that almost all of 16 transmembrane proteins analysed were expressed at levels of 20–50% of total membrane protein with purification yields of 1–2 mg per litre of culture in a certain

vector system (reviewed in Wang *et al.*, 2003). However, several of the proteins that were expressed using the tac promoter did not express with a T7 promoter. The switch from a T7 promoter bearing vector (pETBlue system) to a T5 promoter system (QIAexpress) did not enhance the expression of recombinant proteins in the present study. However, there are several other potential reasons for the failure of target gene expression. Firstly, the target protein itself may interfere with gene expression or with the integrity of the cell (Novagen, 2004). Secondly, a factor that appears to influence target protein stability is the amino acid residue immediately following the N-terminal methionine (the so called penultimate amino acid). The amino acid at this position determines the removal of N-terminal fMet. Processing is catalysed by methionyl aminopeptidase and the degree of removal decreases as the size of the penultimate amino acid side chain increases (Hirel *et al.*, 1989; Lathrop *et al.*, 1992; Novagen, 2004). Only little or no processing was observed when the following amino acids occupied the penultimate position: His, Gln, Glu, Phe, Met, Lys, Tyr, Trp, Arg. However, processing ranged from 16%–97% when any other amino acid occupied this position (Hirel *et al.*, 1989; Lathrop *et al.*, 1992). Thirdly, it was observed that the life span of a protein in *E. coli* is determined by its amino-terminal (N-terminal) amino acid (N-end rule). Tobias *et al.* (1991) reported protein half-lives of only 2 min with amino-terminal Arg, Lys, Phe, Leu, Trp, and Tyr present, whereas all other amino-terminal residues conferred >10 h half-lives on the same protein. The most likely reason for failed target gene expression in the present study is a combination of potential exposure to fMet processing followed by rapid degradation of the polypeptide (N-end rule), as the PnVPP fragments contained Phe (153 bp insert) and Tyr (453 bp insert) in the penultimate position. However, the program ProtParam (http://ca.expasy.org/tools/protparam.html; Walker, 2005), which relies in its half-life predictions on the N-end rule, classified both proteins as stable with a half-life of over >10 h in *E. coli in vivo*.

As the overexpression of the PnVPP protein was unsuccessful, an anti-peptide approach was chosen for the generation of antibodies. Peptides were carefully selected, taking into consideration a number of factors deemed to be important for successful production of anti-peptide antibodies. The number of potential regions was very limited in PnVPP, as the protein is highly hydrophobic and contains many transmembrane loops. Interestingly, the two peptides chosen in this study for immunisation of rabbits and mice correspond to the regions used by Kim *et al.* (1994a) for the production of polyclonal

antibodies. Pep-PP1, which encompasses the Dx_7KxE motif corresponds to PAB_{TK}, and Pep-PP2, situated near the N-terminus and containing the second acidic motif Dx_3Dx_3D, corresponds to PAB_{HK}. Both antibodies, PAB_{TK} and PAB_{HK}, are well characterised and have been used in a large number of studies (e.g. Drozdowicz *et al.*, 2000; Marchesini *et al.*, 2000; Ruiz *et al.*, 2001a; Martinez *et al.*, 2002; Drozdowicz *et al.*, 2003).

The immune response in rabbits was variable for both peptides. Only rabbit #2, inoculated with peptide Pep-PP1, showed a good response in ELISAs. Monoclonal antibodies were obtained from a mouse immunised with Pep-PP2. In the initial screen which used immunofluorescence assays, 11 out of 709 hybridoma cell lines (1.5%) reacted positively with *P. nicotianae* zoospores. However, this number dropped to only four positives clones after further rounds of screening (0.6%). This percentage of positive clones was low when compared to other fusions carried out in our laboratory, although a fusion with similar efficiency was reported with only nine positives out of 721 screened cell lines (1.2%) (Robold and Hardham, 1998). However, there have also been reports of highly successful fusions with 13% and 25% positive clones, respectively (Hardham *et al.*, 1991; Robold and Hardham, 2004). The success of these highly efficient fusions could be due to the fact that whole proteins or cellular fractions of *Phytophthora* species were used for immunisation, offering a greater range of antigens to the immune system, compared to the short peptides used in this study.

P. nicotianae zoospore protein immunoblots labelled with PnVPP antibodies generated in the present study from rabbit #2 reacted with two bands at ~36/45 kDa and ~65/75 kDa, depending on the molecular weight marker used. The occurrence of a lower band around 37-45 kDa in immunoblots of V-PPases has been reported in *Z. mays* and *Acer pseudoplatanus* (reviewed in Rea and Poole, 1993) and is most likely explained by proteolysis of the protein. The inferred molecular weight of the PnVPP protein is 84 kDa, therefore the predicted molecular weight is approximately 10-20 kDa above the observed (higher) band. Inconsistencies between the apparent and predicted molecular weight of V-PPases on SDS-PAGE are reported frequently (e.g. Nyren *et al.*, 1991; Drozdowicz *et al.*, 1999) and are most likely due to the high hydrophobicity of the protein and incomplete saturation by SDS (Maddy, 1976; Nyren *et al.*, 1991).

Peptide competition assays were carried out in order to determine the specificity of the antibodies that were produced. Initial experiments using *P. nicotianae* zoospore extracts

did not show a reduction in labelling intensity when antibodies were pre-incubated with their respective immunising peptide. Immunodot blots showed, however, that antibodies were indeed able to recognise their peptides and labelling was abolished using a 500-fold excess of the peptide. An immunoblot with protein extracts from all four developmental stages of *P. nicotianae* showed consistent labelling of both bands in all stages, however, none of the bands were blocked by peptide competition. The presence of the two antigenic bands in all developmental stages is in contrast to the gene expression profile obtained from northern blots in which expression was high in zoospores and germinated cysts, moderate in vegetative hyphae and not detectable in sporulating hyphae.

Taken together, these results indicate that the labelling of antibodies from rabbit #2 in immunoblots is non-specific.

Only two out of the four monoclonal antibodies reacted in immunoblots. Antibody 5D4 labelled a range of bands and was therefore not used in further studies. 2G8 labelled a double band just below 50 kDa in the initial experiments. However, the use of freshly prepared *P. nicotianae* zoospore extract in immunoblots shifted the labelled band to ~75 kDa, with a faint ~50 kDa band still observable. This supports the idea of proteolytic degradation of the protein. However, when compared on the same immunoblot, polyclonal antibody from rabbit #2 and monoclonal antibody 2G8 did not label corresponding bands. So while it is intriguing that the polyclonal and monoclonal antibodies bind to polypeptides of approximately the right molecular weight, this observation, together with the fact that no decrease in labelling intensity was observed in peptide competition assays with antibody 2G8, suggests that labelling by the antibodies is non-specific in immunoblots.

Immunolocalisation studies using the antibodies generated in the present study on zoospores from *P. nicotianae* fixed in formaldehyde plus glutaraldehyde showed labelling of some of the WEVs and the plasma membrane. There was no consistent labelling observed in zoospores fixed in formaldehyde alone. The undefined labelling pattern is most likely due to poor preservation of the membrane, as fixation in formaldehyde alone (i.e. without glutaraldehyde) causes degradation of the plasma membrane. The fact that only some WEVs were labelled in zoospores prepared in formaldehyde plus glutaraldehyde fixation is consistent with the fact that the bladder needs to be open to the external environment for labelling to take place. As this only occurs during the water

expulsion phase of the WEV cycle, it would explain why some WEVs visible in bright field were not labelled by the antibody (Mitchell and Hardham, 1999). However, as both monoclonal and polyclonal antibodies were deemed non-specific in immunoblots, localisation patterns obtained in immunofluorescence experiments have to be viewed with care as they might also be a result of non-specific binding of the antibodies.

3.3.4 Is PnVPP involved in osmoregulation?

While immunoblot experiments throw doubt on the specificity of the polyclonal and monoclonal antibodies for the *P. nicotianae* V-PPase, the immunolocalisation of the enzyme to the zoospores' plasma membrane and bladder of the WEV would fit with the localisation pattern of this protein in other organisms. Concurrent vacuole and plasma membrane localisation has been observed in a range of organisms including plants, protozoa and algae, for example *Pisum sativum* (Robinson *et al.*, 1996), *C. reinhardtii* (Robinson *et al.*, 1998), *P. falciparum* (Luo *et al.*, 1999), *T. cruzi* (Scott *et al.*, 1998), *T. gondii* (Rodrigues *et al.*, 2000) and *L. wallacei* (Moraes Moreira *et al.*, 2005). However, as stated above, non-specific binding of antibodies generated in the present study can not be entirely excluded, and thus interpretation of result should be viewed with care.

CVs, including the WEV in *Phytophthora* species, are thought to be composed of two membrane systems: a) the spongiome, containing a V-ATPase and b) the bladder which lacks the V-ATPase enzyme but contains a P-type ATPase instead (Allen and Naitoh, 2002). An immunolocalisation study in *P. nicotianae* showed indeed that the V-ATPase is confined to the spongiome, whereas a P-type ATPase is present in the bladder and plasma membrane (Mitchell and Hardham, 1999). As the antibody used in the present study labelled the WEV only in cells fixed in formaldehyde plus glutaraldehyde, it is thought that the PnVPP protein is also localised to the bladder but not the spongiome and therefore co-localises with the P-type ATPase but not the V-ATPase. In contrast, in a series of studies on three organisms bearing CVs (*D. discoideum*, *C. reinhardtii* and *T. cruzi*) conducted in one laboratory, a co-localisation of the V-ATPase and V-PPase to the bladder was claimed (Ruiz *et al.*, 2001a; Marchesini *et al.*, 2002; Rohloff *et al.*, 2004). However, as results were obtained by immunolocalisation and biochemical studies on sub-cellular fractions after gradient centrifugation only, convincing evidence (i.e. studies at the ultrastructural level with the V-ATPase antibodies) for a co-localisation of these two enzymes is lacking. Contamination of the bladder fraction with other membrane

fractions and/or co-sedimentation with the spongiome fraction during the fractionation process is highly likely and would explain activity of spongiome-specific enzymes that were observed in the bladder fractions. Furthermore, occurrence of the V-ATPases in the spongiome has been shown clearly by immunolocalisation studies with monoclonal antibodies using electron microscopy in several organisms (reviewed in Allen, 2000 and Mitchell and Hardham, 1999; Allen and Naitoh, 2002), and there is no evidence to date that the V-PPase protein co-localises with the V-ATPase in the spongiome.

Co-localisation of V-PPases and V-ATPases has, however, been clearly demonstrated for plant vacuoles (reviewed in Barkla and Pantoja, 1996; Maeshima, 2001), acidocalcisomes of apicomplexans and trypanosomatids (reviewed in Docampo and Moreno, 2001) and the 'mass-dense' granules from *D. discoideum* (Marchesini *et al.*, 2002). In these organelles, V-PPases function most likely as back-up and/or adjunct H^+-pumps for the V-ATPase. Furthermore, it has been proposed that acidocalcisomes are linked to the CV function and may be involved in osmoregulation. The exact mechanism by which CV-containing cells expel water remains elusive to date. Cho and Fuller (1989) proposed the following working model for the expulsion of water by the WEV in *P. palmivora*: the surrounding spongiome develops into the bladder, and upon water expulsion the bladder fuses with the plasma membrane. Similar models have been suggested for other organisms, such as *D. discoideum* (reviewed in Allen, 2000; Allen and Naitoh, 2002).

As the PnVPP protein is presumably located in the bladder compartment of the CV complex in *P. nicotianae* and not in the spongiome, a function as a back-up H^+-pump for the V-ATPase, as reported for acidocalcisomes and plant vacuoles, and thus involvement in WEV function in *P. nicotianae* zoospores, seems unlikely. This notion is supported by the results of *in vivo* inhibition experiments using *P. nicotianae* zoospores with the V-PPase-specific inhibitor 1,1-bisphosphonate compound AMBP. Although pyrophosphatase activity assays of microsomal fractions of *P. nicotianae* zoospores showed PP_i hydrolysis at a rate of approximately 0.1 µmol PP_i/min/mg protein, which is comparable to that of most reported endogenous V-PPase activities found in membrane preparations from several organisms (range from 0.02 – 1.1 µmol PP_i/min/mg - Maeshima and Yoshida, 1989; Scott *et al.*, 1998; Marchesini *et al.*, 2000; McIntosh *et al.*, 2001; Martinez *et al.*, 2002), and inhibition by AMBP reduced PP_i hydrolysis by approximately 60% at a concentration of 30 µM, which is within the lower range of

reduction as demonstrated by other groups (Rodrigues *et al.*, 1999a; McIntosh *et al.*, 2001; Marchesini *et al.*, 2002), *in vivo* experiments using up to 10 µM AMBP failed to show a reduction in WEV cycle time in *P. nicotianae* zoospores. However, it was reported that the efficacy of AMBP *in vivo* is usually less than its efficacy *in vitro*, which is most likely due to the bulky anionic character of AMBP and its slow permeation of membranes (Drozdowicz *et al.*, 2003). In *P. falciparum* cultures, for example, parasite growth was diminished by only 20% upon the addition of 250 µM AMBP, and concentrations of 500–1000 µM were required to reduce growth by 85% (McIntosh *et al.*, 2001). As only concentrations up to 10 µM AMBP were tested in *in vivo* experiments in the present study, a definite conclusion about the impact on AMBP on PnVPP function cannot be drawn at this point. Further experiments using higher concentrations of AMBP need to be performed, in order to fully clarify the possible effect of AMBP on PnVPP *in vivo*.

3.3.5 Gene silencing

Gene silencing is an important tool for the analysis of gene function in today's research. Of particular interest is the post-transcriptional gene silencing approach of RNAi, which employs the introduction of short interfering sequence-specific dsRNA molecules which degrade endogenous mRNAs. In this study, a dsRNA construct carrying a 303 bp fragment of the *PnVPP* gene has been cloned successfully into the *Phytophthora* transformation vector pTH210. Unfortunately, it was not feasible to perform the transformation experiments due to time constraints and the fact that protocols for stable transformation of *P. nicotianae* were still being developed during the course of this study.

Gene silencing of the PnVPP protein and the analysis of transformants might be a challenging undertaking. In order to clarify the potential role of PnVPP in WEV function and osmoregulation, zoospores are the developmental stage of interest. The protein has been shown to localise not only to the WEV but also to the plasma membrane in zoospores, indicating that it might be involved in other, potentially vital cell processes. Therefore, silencing of the gene might be fatal in this or other developmental stages. Growth of mycelium might be completely abolished or significantly impaired, thus not giving rise to the required developmental stage. Furthermore, it is also possible that zoospores are not viable due to the loss of V-PPase activity.

In conclusion, this study presents information on the gene structure and gene expression of PnVPP from *P. nicotianae*. The protein is predicted to be a Type II H^+-translocating pyrophosphatase similar to the proteins of *A. thaliana* (AVP2) and *P. falciparum* (PfVPP2), for example. The exact role of PnVPP in *P. nicotianae* has not been determined and further studies using, for example, dsRNA gene silencing of PnVPP should help to elucidate the function of the protein *in vivo*.

Chapter 4

Aquaporins in *P. nicotianae*

Chapter 4: Aquaporins in *P. nicotianae*

4.1 Introduction

4.1.1 Aquaporins – An overview

As outlined in Chapter 3, transport of water and solutes across membranes can occur in a number of ways. As diffusion of water through membranes is a slow process, the existence of proteins that form water-specific membrane channels, which allow for rapid water permeation across membranes, was proposed a long time ago. However, it was only in 1986 that Benga and colleagues confirmed the presence of a water channel in human erythrocytes (Benga *et al.*, 1986a; Benga *et al.*, 1986b). Two years later, the first channel-forming integral membrane protein (CHIP) was isolated from human erythrocytes and partially characterised (Denker *et al.*, 1988). However, a connection to the long sought-after water channels was only made some time later. Over the following years, the 28 kDa protein, originally named CHIP28, was extensively analysed on a molecular level and in functional studies, and the first 'trademarks' of this new protein family were manifested (reviewed in Agre *et al.*, 1993a; Matsuzaki *et al.*, 2002; Benga, 2003). In 1993, after the discovery of two more water channel proteins, CHIPs were renamed to aquaporins (Agre *et al.*, 1993b). Since then, a vast number of aquaporins and closely related MIPs (major intrinsic proteins) have been discovered in plants, microbes and mammals. In a recent phylogeny study, Zardoya compiled a list of more than 450 non-redundant protein sequences of MIPs from NCBI databases (Zardoya, 2005). Furthermore, a newly created database, MIP*DB* (El Karkouri *et al.*, 2005), integrates 785 MIPs from more than 200 organisms, emphasising the widespread distribution of these proteins and their abundance.

MIPs are historically classified into aquaporins and glycerol facilitator-like proteins (GLPs), based on their selectivity for solute transport. Aquaporins are only permeated by water, whereas GLPs also allow transport of glycerol and sometimes other small molecules such as urea (reviewed in e.g. Zardoya, 2005; Kruse *et al.*, 2006). More recently, however, it has become clear that functions of proteins from the MIP family are more diverse than previously thought. Some proteins feature water permeation only, whereas others are specific for glycerol transport. Yet others are permeable to both water and glycerol (and other solutes), and were therefore named aquaglyceroporins. Additionally,

one protein (AQP6) was reported to have only minimal water transfer capability, but transfers anions efficiently (Yasui *et al.*, 1999; Ikeda *et al.*, 2002a). As the categorisation of MIPs into only two main groups seemed inappropriate in view of the large array of functions and distributions for these proteins, El Karkouri and colleagues proposed a more precise classification of the MIP family (El Karkouri *et al.*, 2005). Figure 4.1 shows the eight proposed functional/biological groups: PIP (plasma membrane intrinsic proteins), AQPe (aquaporins of eukaryota), AQPp (aquaporins of prokaryota), TIP (tonoplast intrinsic proteins), NIP (NOD26-like intrinsic proteins), GLPp (GLPs of prokaryota), GLAe (aquaglyceroporins (GLA) of eukaryota) and GLAp (GLAs of prokaryota). As this new nomenclature has been proposed very recently, it has not been fully adopted by the scientific community yet, with the majority of publications still referring to the 'old' two cluster system, AQP and GLP.

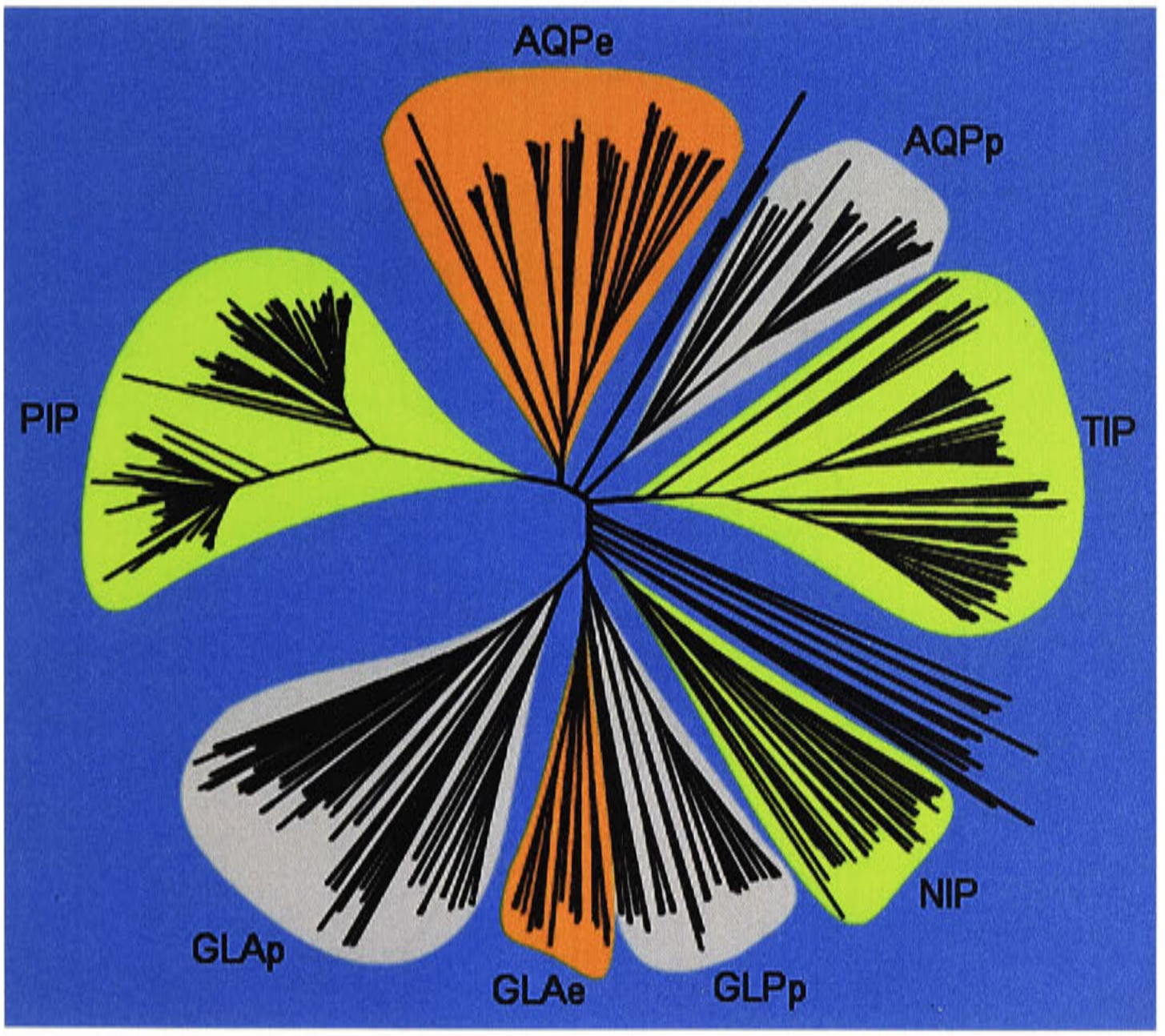


Figure 4.1 A sequence comparison of MIP proteins, showing the eight functional/biological groups as proposed by El Karkouri *et al.* (2005). Sequences corresponding to uncoloured branches were not included in the sequence analysis. For details on phylogenetic analysis see El Karkouri *et al.* (2005).

4.1.2 Aquaporin structure – the hourglass model

Proteins of the MIP family are composed of a single polypeptide chain of approximately 270 aa. Hydropathy plot analyses and transmembrane prediction programs suggest six transmembrane helices (1-6), which are connected by five loops (A-E). Loops A, C and E are extracellular and loops B and D are intracellular, locating both the N-terminal and C-terminal end of the protein to the cytoplasm. The two halves of the protein are sequence-related and probably evolved by gene duplication. The two connecting loops B and E are highly conserved and contain the consensus motif, asparagine-proline-alanine (NPA). These loops are considered to fold back into the membrane from opposite sides (i.e. loop B entering from the cytoplasm and loop E from the extracellular side), where they overlap and form a narrow, aqueous pore (Figure 4.2, reviewed in Matsuzaki *et al.*, 2002; Agre, 2004; King *et al.*, 2004; Takata *et al.*, 2004; Zardoya, 2005; Kruse *et al.*, 2006).

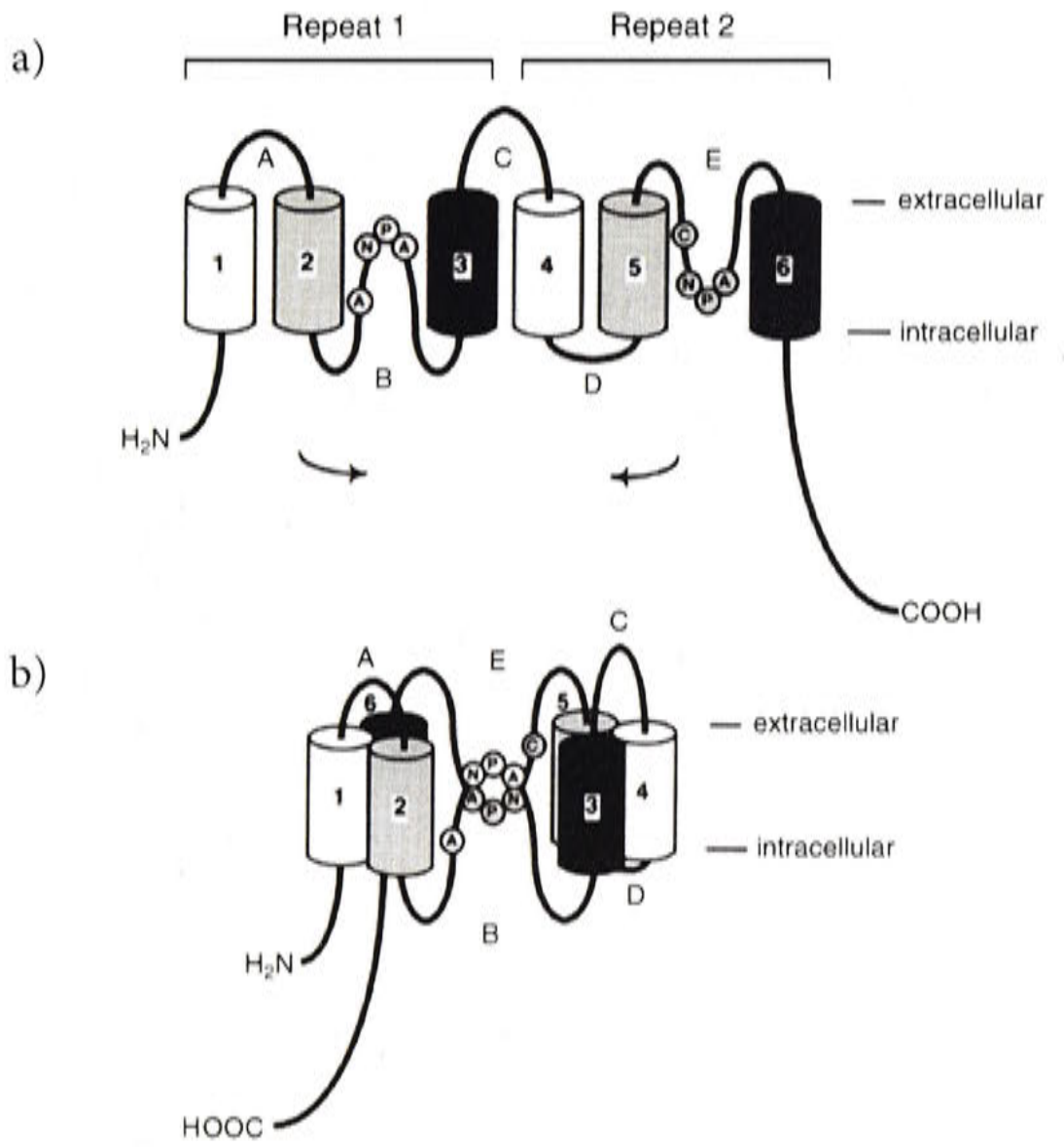


Figure 4.2 Hourglass model for the membrane topology of the AQP1 subunit. a) Schematic representation of the folding of loops B and E; the loops overlap within the lipid bilayer to form a single aqueous pathway, b) Three-dimensional structure of the AQP1 subunit. (Source: Jung *et al.*, 1994; Agre, 2004)

The transmembrane orientation of the AQP1 molecule was initially confirmed using antibodies to both protein termini and the loop domains. The so-called 'hourglass-model' was later verified by several groups using X-ray crystallography and cryo-electron microscopy. Table 4.1 shows on overview of 2D and 3D mapped MIP proteins characterised to date (summarised in Agre, 2004; King *et al.*, 2004; Takata *et al.*, 2004; Wang *et al.*, 2005; Tornroth-Horsefield *et al.*, 2006).

Table 4.1 Overview of high-resolution 3D structure models for various MIPs. (Res. – Resolution, X-ray – X-ray crystallography of 3D crystals; EM – cryo-electron microscopic examination of 2D crystals)

MIP	Origin	Res.	Method	Reference
Mammals				
AQP0	Bovine lens	2.2 Å	X-ray	Harries *et al.*, 2004
	Sheep lens	3.0 Å	EM	Gonen *et al.*, 2004
AQP1	Human erythrocyte	7.0 Å	EM	Cheng *et al.*, 1997
		6.0 Å	EM	Walz *et al.*, 1997
		3.8 Å	EM	Murata *et al.*, 2000
		3.7 Å	EM	Ren *et al.*, 2001
	Bovine erythrocyte	2.2 Å	X-ray	Sui *et al.*, 2001
AQP2	Human (cDNA)	4.5 Å	EM	Schenk *et al.*, 2005
Bacteria				
GlpF	*Escherichia coli* (cDNA)	3.7 Å	EM	Braun *et al.*, 2000
		2.2 Å	X-ray	Fu *et al.*, 2000
AqpZ	*Escherichia coli* (cDNA)	8.0 Å	EM	Ringler *et al.*, 1999
		2.5 Å	X-ray	Savage *et al.*, 2003
AqpM	*Methanothermobacter marburgensis* (cDNA)	1.7 Å	X-ray	Lee *et al.*, 2005
Plants				
α-TIP	*Phaseolus vulgaris*	7.7 Å	EM	Daniels *et al.*, 1999
SoPIP2;1	*Spinacia oleracea* (cDNA)	5.0 Å	EM	Kukulski *et al.*, 2005
		2.1 Å	X-ray	Tornroth-Horsefield *et al.*, 2006

It has been shown that the monomeric units of MIP proteins associate into a tetramer in the cell membrane, thus forming a fifth pore in the centre (reviewed in Agre and Kozono, 2003; King *et al.*, 2004; Kruse *et al.*, 2006; Figure 4.3). Each monomer contains an independently functioning pore facilitating the transport of solutes across membranes, whereas the channel at the centre of the tetramer does not form a path for water (Murata *et al.*, 2000). Recently, it has been proposed that this channel might conduct ions such as K^+, Cs^+, Na^+ and tetramethylammonium (Saparov *et al.*, 2001; Yool and Weinstein, 2002).

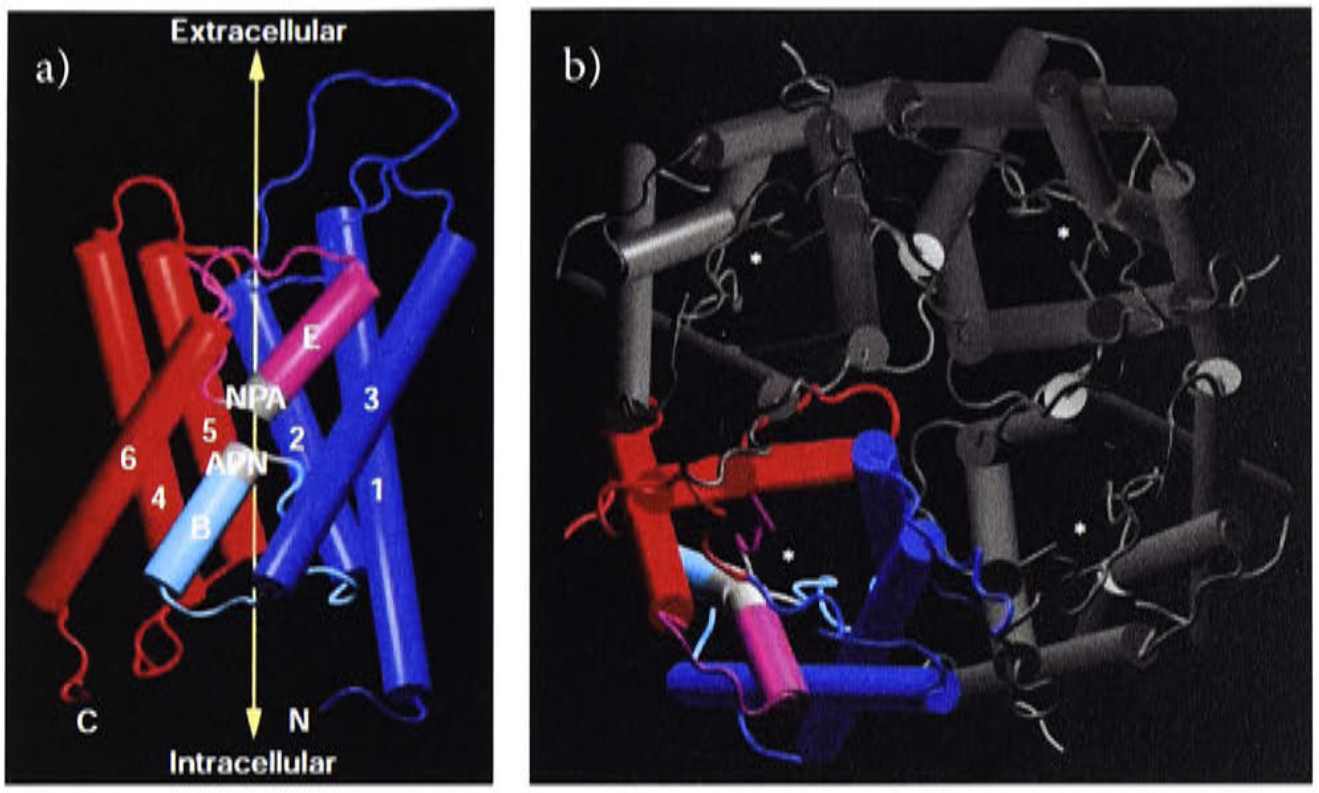


Figure 4.3 The structure of AQP1. a) The structure of the AQP1 monomer is shown, with membrane-spanning helices numbered 1–6 and displayed as rods. The N-terminal half of the molecule is shown in purple and light blue, and the C-terminal half is shown in red and pink. Loops B and E, which fold into the membrane to form the pore as well as the conserved NPA motifs are labelled. Parts of loops B and E that form α-helices are shown as rods. The arrow highlights the route taken by water, which can move in both directions through the channel. b) The AQP1 tetramer, as seen from above. Asterisks denote the location of the water pore in each subunit. These structures are based on the PDB coordinates 1J4N. (Source: King *et al.*, 2004)

The most thoroughly investigated MIPs regarding structure-function relationships to date are AQP1 (human and bovine), a channel transporting H_2O only, and GlpF from *E. coli*, a glycerol channel. Even though these proteins share only ~30% sequence identity at an amino acid level, their 3D-structure is remarkably similar (Unger, 2000). Comparison of the two 3D-models as well as a number of molecular dynamics simulations gave insight into the transport process of water or glycerol through the protein's pore (Unger, 2000; de Groot and Grubmuller, 2001; Jensen *et al.*, 2001; Sui *et al.*, 2001; Fujiyoshi *et al.*, 2002; Stroud *et al.*, 2003; Roux and Schulten, 2004; Vidossich *et al.*, 2004; Wang *et al.*, 2005). In AQP1, the narrowest part of the pore was reported to be 3Å in diameter, permitting only water molecules (2.8 Å in size) but not ions or larger solutes to pass (Murata *et al.*, 2000). The narrowest section of GlpF, on the other hand, is approximately 4 Å, which is wide enough to allow glycerol and possibly other small solutes to permeate the pore (Fu *et al.*, 2000). The narrowest constriction of the protein, the 'selectivity filter', is shaped by four residues. Figure 4.4 shows an overlay of the selectivity filter regions of bovine AQP1 and *E. coli* GlpF with the four residues of interest denoted. In Table 4.2 a comparison of these residues, from organisms that have

been characterised by X-ray crystallography to date, is shown. Three of the four selectivity filter residues are highly conserved in orthodox aquaporins (Phe, His and Arg). Substitution of the histidine and cysteine residues in AQP1 with glycine and phenylalanine, respectively, in GlpF from *E. coli* renders the protein capable of glycerol transport due to a change in the protein's architecture and polarity (Sui *et al.*, 2001). In general, selectivity of the MIP channels seems to be highly dependent on the architecture of the protein (i.e. steric organisation of pore-facing side chains), the narrowest point of the selectivity filter and the hydrophobicity and polarity of the pore.

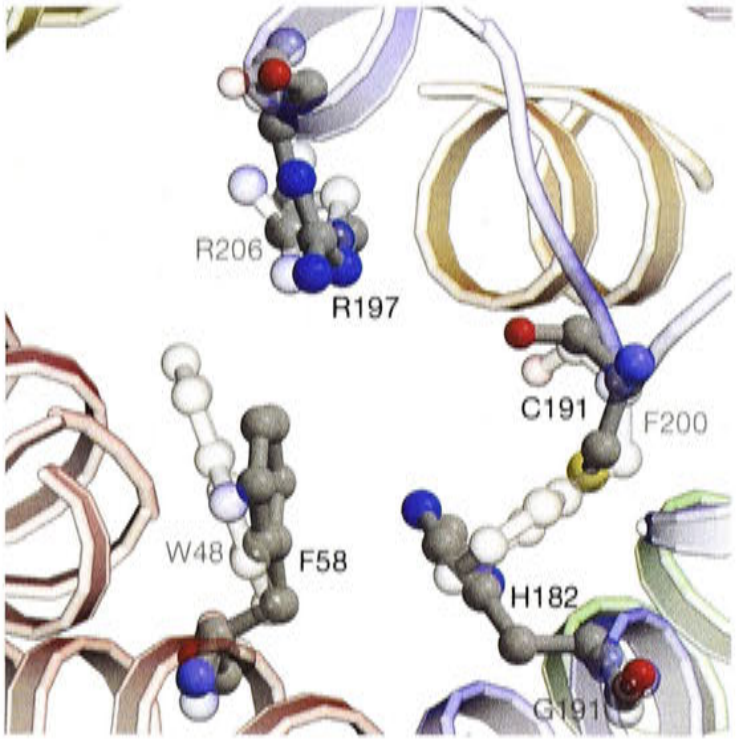


Figure 4.4 Residues involved in the formation of the bovine AQP1 constriction region (H182, R197, F58 and C191) are depicted in solid colours while the transparent side chains are those of the equivalent residues found in the structure of GlpF (G191, R206, W48 and F200). These side chain differences result in a larger and more hydrophobic constriction region in GlpF. (Source: Sui *et al.*, 2001)

Table 4.2 Selectivity filter residue comparison

MIP	Organism	Reference	Position ①	②	③	④	∅
AQP1	Human	Murata *et al.*, 2000	F56	H180	R195	C189	3 Å
	Bovine	Sui *et al.*, 2001	F58	H182	R197	C191	2.8 Å
AQP0	Sheep	Gonen *et al.*, 2004	F48	H172	R187	A181	~2 Å
AQPZ	*E. coli*	Savage *et al.*, 2003	F43	H174	R189	T183	~2 Å
AqpM	*M. marburgensis*	Lee *et al.*, 2005	F62	I187	R202	S196	2.5 Å
GlpF	*E. coli*	Fu *et al.*, 2000	W48	G191	R206	F200	3.8 Å

In orthodox aquaporins, water passes through the pore as a single line of water molecules, guided by the hydrophobic surface lining bestowed by several residues within the pore, thus allowing for rapid water transportation (Murata *et al.*, 2000). It was indicated that transport of protonated water (H_3O^+) through the channel is not only

restricted by several features of the protein, but is also energetically highly unfavourable (Sui *et al.*, 2001). Permeation by protons is prevented by three barriers: a) size restriction (see above), b) electrostatic repulsion by a conserved arginine residue at position 195 (human AQP1) situated in loop E, and c) water dipole reorientation by the two highly conserved arginine residues of the NPA motifs, which prevent hydrogen bonding in the single line of water (Figure 4.5, reviewed in Kozono *et al.*, 2002; Agre and Kozono, 2003; Liu *et al.*, 2006).

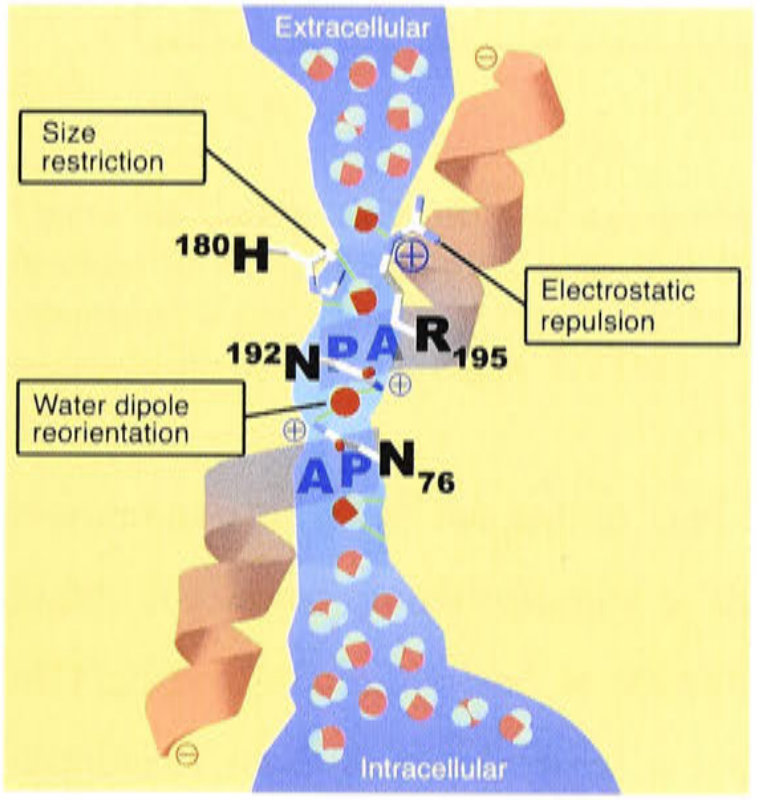
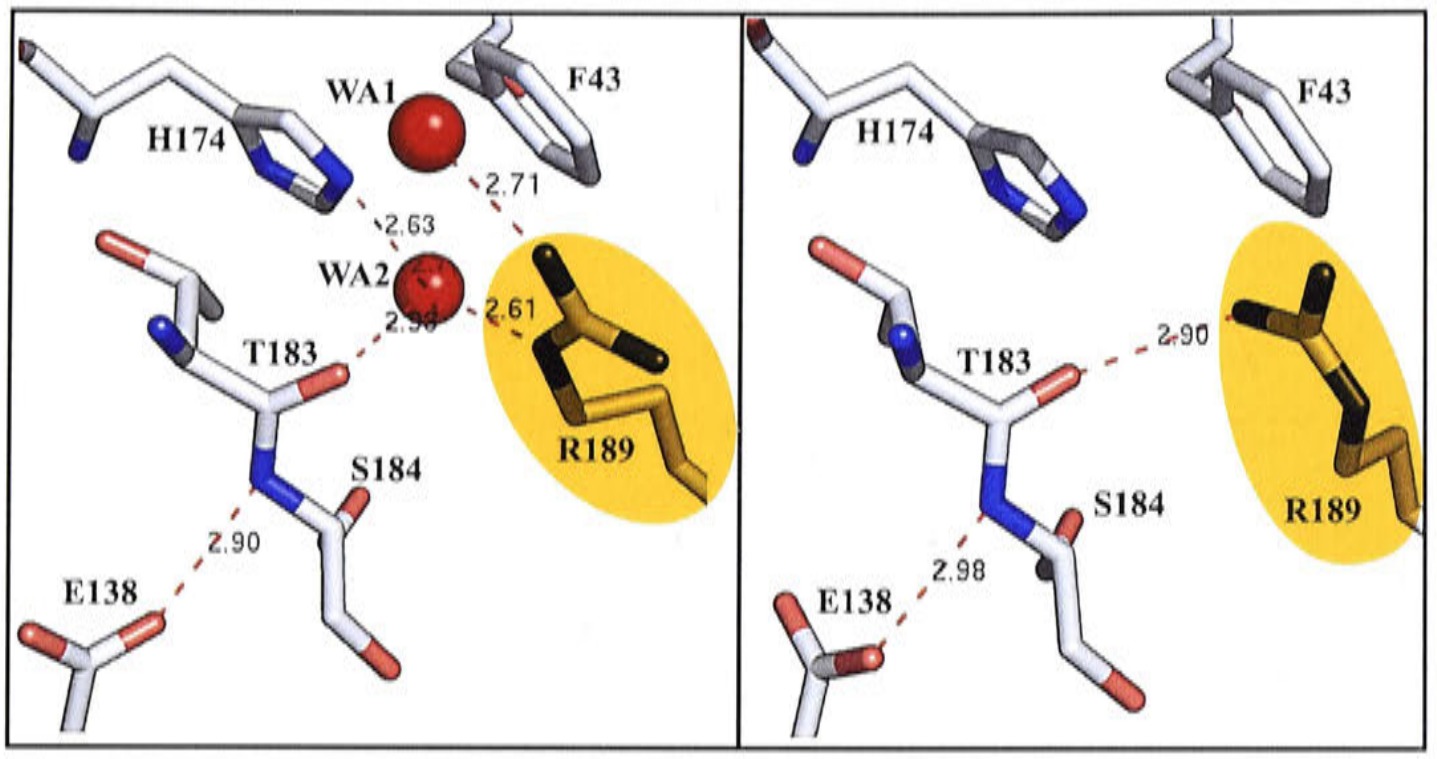


Figure 4.5 Schematic architecture of the channel within an AQP1 subunit. The shape of the aqueous pore (blue) is derived from calculations based on the structure of bovine AQP1. Four water molecules shown in bold colours represent interactions with the pore-lining residues at discrete sites. Bulk water in is depicted in pastels. Three features of the channel specify selectivity for water: (a) Size restriction. 8 Å above the midpoint of the channel, the pore narrows to a diameter of 3 Å. (b) Electrostatic repulsion. A conserved residue (Arg195) at the narrowest constriction of the pore imposes a barrier to cations, including protonated water (H_3O^+). (c) Water dipole reorientation. Two partial helices meet at the midpoint of the channel, providing positively charged dipoles that reorient a water molecule as it traverses this point. (Source: Kozono *et al.*, 2002)

Recently, it has been proposed that aquaporin activity is regulated by the opening and closing of the pore, a process called gating. This mechanism was first observed in molecular dynamics simulations in AqpZ of *E. coli*, where it was suggested that the side chain of Arg189, which is one of the defining residues of the selectivity filter and highly conserved throughout all MIPs, switches between two conformational states, thus opening or blocking the pore to water permeation (Wang *et al.*, 2005). This hypothesis was confirmed by Jiang and colleagues, who presented an X-ray crystallography study on tetrameric AqpZ of *E. coli* (Jiang *et al.*, 2006a). Here it was shown, that aquaporin monomers within a tetramer structure can display two distinct conformations (i.e. fully open or fully closed), based on the orientation of the guanidino group of Arg189 (Figure 4.6).



Open Closed

Figure 4.6 Gating mechanism of aquaporins. The selectivity filter of *E. coli* AqpZ with key residues involved in formation of the constriction point are depicted. Highlighted areas denote the different orientation of the guanidine group of Arg189, responsible for the state of the protein. (WA1/WA2 – water molecule, modified from Jiang *et al.*, 2006a)

Numerous studies at the amino acid level over the years have revealed a number of highly conserved motifs common to the majority of the MIP family members, as shown in Figure 4.7 (summarised in Zardoya, 2005). Furthermore, it was proposed that a number of residues may be used as indicators for channel selectivity and differentiation between aquaporins and GLPs. According to Froger and colleagues, five positions (P1-P5) located in the terminal part of TMD3 and in loop E and TMD6, were clearly identified as residues that are specific for each functional subgroup as determined by sequence comparison and analysis (Table 4.3, Froger *et al.*, 1998). The functional relevance of residues P4 and P5 was shown by substitution of aquaporin to GLP-specific residues in an insect aquaporin, which led to a switch in the channel selectivity conferring glycerol transport properties to the mutant aquaporin protein (Lagree *et al.*, 1999). However, substitutions in these positions in *E. coli* AqpZ and GlpF from aquaporin to GLP-specific residues and *vice versa* did not display a selectivity switch (Borgnia and Agre, 2001). Furthermore, a number of other reports also showed exceptions to this rule predominantly in plant NIPs, all exhibiting residues that lead to mixed assignments (Santoni *et al.*, 2000; Chaumont *et al.*, 2001; Schuurmans *et al.*, 2003). Thus it can be concluded that characterisation and classification of members of this vast protein family cannot easily be achieved solely by analysis of the primary amino acid sequence and the determination of the five positions P1 to P5.

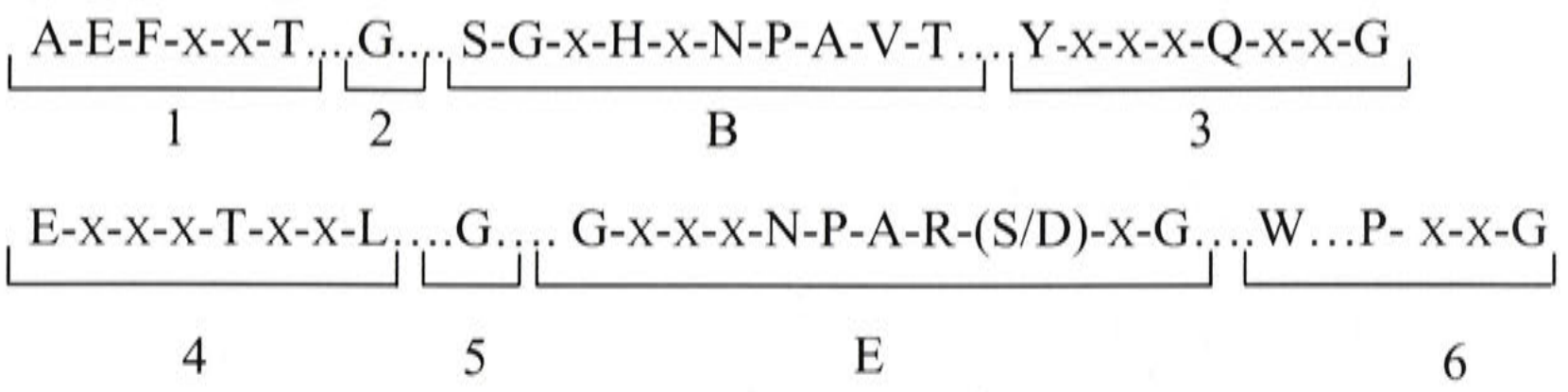


Figure 4.7 Conserved signatures of the MIP molecule that were deduced from comparative analysis of multiple alignments. Numbers indicate TMDs, letters indicate loops. (Source: Zardoya, 2005)

Table 4.3 Differentiation of the two distinct clusters (Aquaporins and GLPs) based on the five positions P1-P5 as proposed by Froger *et al.*, 1998

Residue	Location	Cluster	Differentiation
P1	TMD3/loop C	AQP	non-aromatic
		GLP	aromatic
P2	loop E	AQP	small & uncharged
		GLP	Asp
P3	loop E	AQP	small & uncharged
		GLP	Lys/Arg
P4	TMD 6	AQP	aromatic
		GLP	Pro
P5	TMD 6	AQP	aromatic
		GLP	non-aromatic

4.1.3 Aquaporins in mammals

To date, 13 aquaporins have been reported in mammals. Seven of those are only permeated by water (AQP0, AQP1, AQP2, AQP4, AQP5, AQP6 and AQP8) and fall therefore into the AQP group, whereas the four proteins AQP3, AQP7, AQP9, AQP10 also allow permeation of other small solutes like glycerol and urea and belong to the GLP group (reviewed in King and Yasui, 2002; Takata *et al.*, 2004). Furthermore, these two distinct groups share a similar genomic sequence structure with four exons and six exons, respectively (reviewed in Morishita *et al.*, 2004). They have been mapped to chromosomes 1, 7, 9, 10, 12, 15 and 16, with the genes encoding AQP0, AQP2, AQP5 and AQP6 clustering on chromosome 12 (Kruse *et al.*, 2006).

Two more aquaporin-like genes were identified after the completion of the human genome project by database mining (originally deposited in GenBank as AQPX1 and AQPX2, Ishibashi *et al.*, 2000), and have since been cloned and further analysed (Morishita *et al.*, 2004; Itoh *et al.*, 2005; Morishita *et al.*, 2005). They were renamed AQP11 and AQP12, respectively, and show a distinct genomic sequence structure compared to the other groups, with only three exons. Mapping located them to chromosomes 11 (AQP11) and 2 (AQP12). Both proteins show a divergence in the highly conserved NPA repeats. In contrast to the conventional aquaporins, they have a conserved NPA motif in loop E but variant motifs in loop B with AQP11 displaying asparagine-proline-cysteine (NPC) and AQP12 showing asparagine-proline-threonine (NPT) (Morishita *et al.*, 2004). Furthermore, a highly conserved sequence upstream of the first NPA box (loop B) also differs. Overexpression studies of both proteins in *Xenopus* oocytes showed no localisation to the plasma membrane but instead an intracellular accumulation (Itoh *et al.*, 2005; Morishita *et al.*, 2005). Thus, their function and physiological role remains to be elucidated. An overview of all 13 mammalian aquaporins discovered to date and their solute permeability and potential role is shown in Table 4.4.

Table 4.4 Permeability characteristics and predominant distribution for the known mammalian aquaporin homologues. Solutes displayed in bold indicate high permeability. (adapted from King *et al.*, 2004; Takata *et al.*, 2004)

MIP	Permeability	Tissue distribution	Potential roles
AQP0	Water	Lens	Fluid balance within the lens
AQP1	**Water**	Red blood cell	Osmotic protection
		Kidney	Concentration of urine
		Lung	Alveolar hydration state
		Brain	Production of cerebrospinal fluid
		Eye	Production of tears
AQP2	**Water**	Kidney	Mediates antidiuretic hormone activity
		Vas deferens	Water absorption of luminal fluid
AQP3	**Water** **glycerol** urea	Kidney	Reabsorption of water
		Skin	Maintenance of skin hydration
		Lung	Secretion of water into trachea
		Eye	Production of tears
		Colon	Water absorption
AQP4	**Water**	Brain	CSF fluid balance
		Muscle	Volume change during contraction
		Kidney	Reabsorption of water
		Lung	Bronchial fluid secretion
		Stomach	Transcellular water transfer
AQP5	**Water**	Salivary gland	Production of saliva
		Lacrimal gland	Production of tears
		Sweat gland	Sweat secretion
		Lung	Water transfer
		Cornea	Maintenance of lens transparency
AQP6	Water **anions (NO_3^- > Cl^-)**	Kidney	Glomerular filtration Acid-base metabolism
AQP7	**Water** **glycerol** **urea**, arsenite	Adipose tissue	Transport of glycerol out of adipocytes
		Kidney	Reabsorption of water
		Testis	Volume reduction in dev. spermatids
AQP8	**Water**	Testis	Water secretion
		Kidney	Osmotic equilibrium between cytoplasm and vesicles
		Liver	Secretion of bile
		Pancreas	Secretion of pancreatic juice
AQP9	Water **glycerol** **urea** arsenite	Liver	Transport of glycerol into adipocytes
		Leukocytes	?
		Brain	Production/eeabsorption of CSF
		Testis	Water absorption/secretion
AQP10	Water, **glycerol**, **urea**	Small intestine	Water/small solute absorption
AQP11	?	Testis, Kidney, Liver	?
AQP12	?	Pancreas	?

Aquaporins are thought to play a significant role in human pathophysiology. As each aquaporin shows a unique subcellular, cellular and tissue distribution, with little overlap between the homologues, it is of great importance to elucidate the complex patterns of expression, regulation and membrane targeting of each AQP gene (King *et al.*, 2004). Thus far, there have been only a few reports in which mutations in aquaporins are directly linked to clinical disorders. Mutations of AQP0 have been shown to result in congenital cataract in mice and humans (Shiels and Bassnett, 1996; Berry *et al.*, 2000; Francis *et al.*, 2000). The Colton blood group antigen in human erythrocytes has been reported to be AQP1 (Preston *et al.*, 1994; Smith *et al.*, 1994). Here, a small number of individuals completely lacking AQP1 due to different nonsense or missense mutations (exon deletion or frameshift) resulting in non-functional AQP1 protein, surprisingly did not suffer any apparent clinical consequences. However, erythrocytes exhibited very low water permeability (Preston *et al.*, 1994), and individuals with AQP1 deficiency were later found to have impaired urinary concentrating ability (King *et al.*, 2001). Mutations in human AQP2 result in hereditary (autosomal recessive and dominant), nephrogenic diabetes insipidus (e.g. Deen *et al.*, 1994), a disorder characterised by the kidney's inability to concentrate urine due to the insensitivity to the antidiuretic hormone arginine vasopressin. Involvement of AQP5 in individuals with Sjogren's syndrome, which suffer from dry mouth and dry eyes due to deficient secretion of saliva and tears caused by an autoimmune destruction of salivary glands, has been proposed by several groups (Steinfeld *et al.*, 2001; Tsubota *et al.*, 2001). However, these results remain controversial to date and it is not clear whether AQP5 indeed contributes to the pathogenesis of this disease (Takata *et al.*, 2004).

4.1.4 Aquaporins in plants

Aquaporins in plants are thought to play a major role in the plant's water relations during growth, development and stress responses, with gene expression of homologues in a wide range of tissues and cell types (Maurel, 1997; Tyerman *et al.*, 1999; Maeshima, 2001). In comparison to the mammalian aquaporin family, aquaporins in plants are more abundant, forming large families with up to 35 members and show a greater diversity. In *A. thaliana*, 35 MIP members have been identified by genome analysis (Johanson *et al.*, 2001; Quigley *et al.*, 2001). In *Z. mays* and *O. sativa*, 33 MIP-like isoforms were detected (Chaumont *et al.*, 2001; Sakurai *et al.*, 2005). Plant MIPs are subdivided into four groups based on sequence identity and their localisation within the plant: PIPs, TIPs, NIPs, SIPs (small basic intrinsic proteins). The PIP group can moreover be divided into the two subgroups PIP1 and PIP2 based on sequence divergence, with the PIP1 proteins showing a longer N-terminus, and different water permeability characteristics (reviewed in Alexandersson *et al.*, 2005; Zardoya, 2005; Kaldenhoff and Fischer, 2006b; Kobae *et al.*, 2006). TIPs and PIPs are predominantly located in the vacuolar (tonoplast) and plasma membrane of plants, respectively. Analysis of TIP members by overexpression in *Xenopus* oocytes and complementation of yeast mutants indicated that some TIP isoforms are not only permeable to water, but also to urea, glycerol and NH_4^+/NH_3 (reviewed in Alexandersson *et al.*, 2005; Kaldenhoff and Fischer, 2006b). PIP1 isoforms were shown to have very low water channel activity, but transport other solutes such as glycerol, boric acid, urea and CO_2. Members of the PIP2 subgroup on the other hand, exhibit moderate to strong water channel activity and it was only recently proposed that they might also transport CO_2 (reviewed in Alexandersson *et al.*, 2005; Kaldenhoff and Fischer, 2006b). A member of the NIP family was the first described aquaporin in plants. The prototype nodulin-26 (NOD-26) was isolated from soybean and is located in the peribacteroid membrane of soybean root nodule cells (Fortin *et al.*, 1987). NIPs are thought to be permeable to water, NH_3 and other small solutes like glycerol. The small subfamily of SIPs, originally identified by database mining and phylogenetic analysis (Chaumont *et al.*, 2001; Johanson and Gustavsson, 2002), is the least characterised group of plant aquaporins to date. SIPs are highly basic proteins and show a high sequence divergence when compared to members of the other plant subgroups and even within the subfamily. They exhibit a short N-terminal region and display highly divergent

sequences in the functionally important loop B of the protein. The NPA motif is altered, with replacement of the third position by threonine or leucine in *Z. mays*, and threonine, cysteine or leucine in *A. thaliana* (Chaumont *et al.*, 2001). Recently, three members of the *A. thaliana* SIP family have been analysed by expression studies in yeast and GFP-localisation. The proteins were found in the endoplasmic reticulum (ER) membrane of all tissues except for dry seeds, and water-channel activity was shown for members SIP1;1 and SIP1;2, but not for SIP2;1 (Ishikawa *et al.*, 2005). The physiological function of these aquaporins remains to be elucidated.

Table 4.5 Properties of members of higher plant aquaporin subfamilies. (adapted from Kaldenhoff and Fischer, 2006a)

	Localisation		Substrate	
Subfamily	**Tissue**	**Subcellular**	**specificity**	**Regulation**
TIP	Seed, leaf, root	Vacuolar membranes (tonoplast)	H_2O, urea, NH_3/NH_4^+	Phosphorylation
NIP	Root nodules	Peribacteroid membrane	H_2O, glycerol	Phosphorylation Protein-protein interaction
SIP	Roots, shoots, leafs, flowers, stem	ER membrane	H_2O	unknown
PIP	Roots, shoots, leafs, flowers	Plasma membrane	H_2O, CO_2 glycerol	Heteromerisation of PIP1 and PIP2 monomers, Ca^{2+}, pH, molecular trafficking, phosphorylation

4.1.5 Aquaporins in microorganisms

The physiological role of MIP channel proteins in microbes remains rather elusive to date. Involvement in osmoregulation, freeze tolerance, volume expansion in growing cells and spore formation/germination has been suggested (Calamita, 2000; Hohmann *et al.*, 2000; Hill *et al.*, 2004 Kruse *et al.*, 2006). However, with only a few microbial MIPs functionally characterised (e.g. *E. coli*, *S. cerevisiae*) and the focus of functional analyses predominantly on mammalian and plant MIPs, their function is not conclusively defined. A large number of MIP family members has been identified by database mining of completely sequenced microbial genomes and their functional role has been inferred from sequence comparison and phylogenetic analysis. Tanghe *et al.* (2006) reported, after analysis of all sequenced microbial genomes available in February 2006, a higher occurrence of orthodox aquaporin genes in eukaryotes (67% of 33 species) than in prokaryotes (26% of 193 species). Furthermore, only three out of 22 archaea possess aquaporin-like genes. Interestingly, a large number of microbes lack MIP proteins altogether (Hohmann *et al.*, 2000; Kayingo *et al.*, 2001; Tanghe *et al.*, 2006). Most of these organisms are either animal pathogens or deep-sea inhabitants, and it was hypothesised that, as they might not experience significant changes in osmolarity in their environment, MIP channel activity is not required (Hohmann *et al.*, 2000). However, the existence of an unknown mechanism for water and/or solute transport has not been ruled out to date. Alternatively, sequences of aquaporin-like proteins in those microbes might be so divergent compared to orthodox aquaporins, that identification by similarity searches is unsuccessful. Nevertheless, based on these findings, it was concluded that aquaporins cannot be essential for basic cellular processes and are therefore not required for processes that are important for microbial survival (Tanghe *et al.*, 2006).

The number of genes encoding members of the MIP family in microbes ranges from only one copy to five (Kayingo *et al.*, 2001; Pettersson *et al.*, 2005). Organisms displaying only one gene copy commonly have an aquaglyceroporin instead of an orthodox aquaporin, thus ensuring transport of not only water but also other solutes.

4.1.5.1 Bacteria

The most thoroughly investigated bacterial members of the MIP family are AqpZ and GlpF from *E. coli*. They have been fully characterised on a functional level and crystal structures are available (Fu *et al.*, 2000; Savage *et al.*, 2003). AqpZ is a conventional aquaporin capable of water transport only, whereas GlpF exhibits very low water permeability but also transports glycerol and urea (Maurel *et al.*, 1994). In a recent phylogeny study, Zardoya (2005) analysed 67 eubacterial and four archaean sequences obtained from NCBI databases. Gram-negative and Gram-positive bacteria formed distinct clusters. Moreover, it was shown that GlpF-like proteins were found predominantly in Gram-positive bacteria, whereas Gram-negatives displayed orthodox aquaporins. To date, there have only been three other reports on the functional characterisation of bacterial MIP members other than those of *E. coli*. An aquaporin-like gene, AqpX, was cloned from the Gram-negative bacterium *Brucella abortus* and shown to efficiently transport water but not glycerol, thus classifying it as an orthodox aquaporin (Rodriguez *et al.*, 2000). In 2001, Froger and colleagues cloned the first MIP member of a Gram-positive bacterium. The MIP of *Lactococcus lactis* was shown to belong to the group of aquaglyceroporins, displaying permeability to both water and glycerol (Froger *et al.*, 2001). In the latest report, a member of the archaea, *M. marburgensis*, was investigated. AqpM was identified as an aquaglyceroporin with low water and transient low glycerol permeability (Kozono *et al.*, 2003; Lee *et al.*, 2005).

4.1.5.2 Fungi

Until recently, the only functionally characterised fungal MIP members were those of the baker's yeast *S. cerevisiae* (Bonhivers *et al.*, 1998; Laize *et al.*, 1999; Laize *et al.*, 2000; Carbrey *et al.*, 2001; Meyrial *et al.*, 2001; Sidoux-Walter *et al.*, 2004; Karpel and Bisson, 2006). *S. cerevisiae* contains four MIP-like genes (as identified by genome analysis): two orthodox aquaporins (*Aqy1* and *Aqy2*) and the two aquaglyceroporins *Fps1* and *Yfl054* (Pettersson *et al.*, 2005). Sequence analysis of fungal aquaglyceroporins suggested a separation of this group into three subgroups: Fps1-like proteins (contain a conserved regulatory region in the N-terminus), Yfl054-like proteins (very long N-terminal extension with a conserved stretch), and a third divergent group only found in filamentous fungi, containing proteins not falling into either of the first two groups (Pettersson *et al.*, 2005). Database mining and phylogenetic analysis suggests that each fungal genome has one aquaporin and one *GlfP*-like gene, with the possibility of the

existence of a second copy of any of the two groups in some genomes (Zardoya, 2005). However, a different study revealed the existence of some organisms which only possess a single MIP protein, usually an aquaglyceroporin (Pettersson *et al.*, 2005).

Analysis of the two orthodox aquaporins in *S. cerevisiae* led to an interesting finding. *Aqy1* of commonly used laboratory strains exhibits a number of point mutations leading to a frameshift and thus to a non-functional protein (Bonhivers *et al.*, 1998). A survey of 52 yeast strains revealed that only wild-type and industrial strains, and a very small number of laboratory strains, contain a functional *Aqy1* allele (Laize *et al.*, 2000). Laboratory strain *Aqy2* was found to have an 11-bp deletion leading to a premature stop-codon in the middle of the gene, hence expressing a non-functional protein (Carbrey *et al.*, 2001). Of more than 50 yeast strains analysed, only strain Σ1278b (a wild-type strain derived from commercial baker's yeast) and its derivatives was found to contain an intact ORF for the *Aqy2* gene (Laize *et al.*, 2000). Water transport of Aqy1 and Aqy2 from strain Σ1278b, which contains functional alleles for both genes, was demonstrated in the *Xenopus* oocytes expression system (Aqy1 - Bonhivers *et al.*, 1998; Laize *et al.*, 1999; Carbrey *et al.*, 2001) and by stopped-flow analysis of microsomal vesicles (Aqy2 - Meyrial *et al.*, 2001). Interestingly, *AQP* null mutants (e.g. most laboratory strains) do not show any detrimental effects when compared to strains expressing functional aquaporins (Bonhivers *et al.*, 1998; Carbrey *et al.*, 2001). On the contrary, cells lacking aquaporins were shown to exhibit greater viability when exposed to repeated rapid changes in osmolarity, in comparison to the parental wild-type strain (Bonhivers *et al.*, 1998). Other observed phenotypes for aquaporin null mutants are increased cell surface hydrophobicity along with stronger cell aggregation resulting in greater flocculence, and distinct haploid invasive growth (Carbrey *et al.*, 2001). In a more recent study on ten *S. cerevisiae* wine strains, which contain functional Aqy1 with different water transfer capacities and non-functional Aqy2, it was demonstrated that these strains do not require aquaporins to adapt to stress. Aqy1 null mutants did not show decreased spore fitness or impaired fermentation ability upon stress factors, such as increased temperature or nitrogen limitation (Karpel and Bisson, 2006). Taken together, the results of these studies show that the physiological role of aquaporins in yeast still remains somewhat undefined.

Of the aquaglyceroporins, Fps1 of *S. cerevisiae* is the most thoroughly investigated to date. It has been shown to transport a number of solutes other than glycerol, and it has been demonstrated to play a crucial role in osmoregulation and the determination of intracellular glycerol levels (for review and further information see Hohmann *et al.*, 2000; Kayingo *et al.*, 2001; Pettersson *et al.*, 2005).

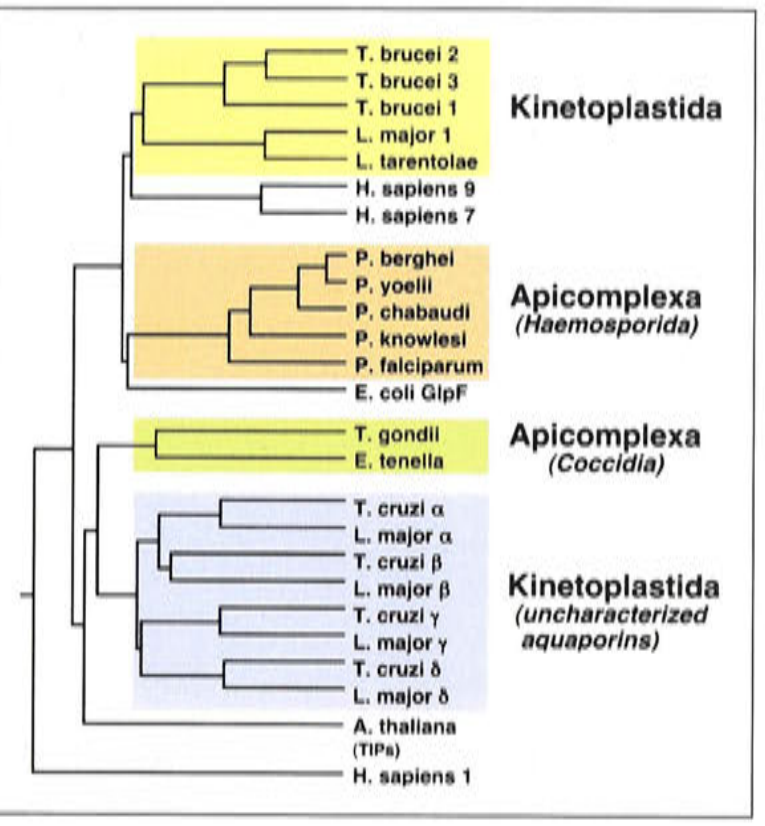
4.1.5.3 Algae and protozoa

Recently, a number of fully sequenced protozoan genomes has become available. Research focuses mainly on protozoan parasites which pose a major threat to humans. Genome analysis showed the existence of at least one MIP-like protein in protozoan parasites with the exception of the apicomplexan *Cryptosporidium parvum* which lacks MIP-like sequences altogether (Beitz, 2005). As a general rule for parasitic protozoans, apicomplexan species appear to express one MIP-like protein only (usually an aquaglyceroporin), whereas members of the class Kinetoplastida show up to five genes (Figure 4.8, summarised in Beitz, 2005). Functional characterisation of aquaporins has been carried out in five of these organisms to date as shown in Table 4.6. Involvement of aquaporins and aquaglyceroporins in osmoregulation and glycerol metabolism, respectively, has been proposed (reviewed in Beitz, 2005). Recently, an aquaporin gene from another obligate intracellular parasitic protozoan of the phylum Microsporidia, which is characterised by the production of resistant spores, was functionally characterised. The genome of *Encephalitozoon cuniculi*, a predominantly mammalian parasite, contains one putative aquaporin sequence and the protein was shown to be highly permeable to water but not to glycerol or other solutes (Ghosh *et al.*, 2005). Germination of microsporidian spores is thought to be dependent on rapid influx of water, thus the discovery of a functional aquaporin in these organisms suggests its involvement in this process (Ghosh *et al.*, 2005).

Aquaporins have furthermore been discovered in *D. discoideum*, where two gene copies (*WacA* and *AqpA*) are present (Flick *et al.*, 1997; Mitra *et al.*, 2000). No functional characterisation in regards to permeability of aquaporins has been performed to date. However, it was shown that disruption of the prespore-specific *AqpA* gene does not affect the organism's growth and morphology, but renders formed spores nonviable (Mitra *et al.*, 2000). The first and to date only report on the isolation and functional characterisation of a MIP-like protein from algae emerged in 2004. Anderca and

colleagues showed that *C. reinhardtii* possesses a functional MIP (CrMIP1), which is capable of glycerol but not water transport. It was proposed that CrMIP1 localises to the CV and plasma membrane of *C. reinhardtii*, where it plays a role in glycerol transport across membranes (Anderca *et al.*, 2004).

Phylum (order)	Species	No. of AQP genes	No. of characterized AQPs
Apicomplexa (Haemosporida)	*P. falciparum*	1	1
	P. berghei	1	–
	P. yoelii	1	–
	P. chabaudi	1	–
	P. knowlesi	1	–
Apicomplexa (Coccidia)	*T. gondii*	1	1
	E. tenella	1	–
	C. parvum	–	–
Kinetoplastida	*T. brucei*	3	3
	T. cruzi	4	1
	L. major	5	1*



a) b)

Figure 4.8 Overview of protozoan aquaporins. (a) Number of identified and experimentally characterized aquaporins in protozoan parasites. (b) Phylogenetic tree of protozoan aquaporins based on protein similarity. For comparison, the human aquaglyceroporins 7 and 9, the human AQP1, the *E. coli* GlpF and the *A. thaliana* γ-TIP1 have been included. (Source: Beitz, 2005)

Table 4.6 Characteristics of functionally analysed apicomplexan and kinetoplastid aquaporins. (nd – not determined)

		Transport of			
Organism	**Name**	**Water**	**Glycerol**	**Urea**	**Other**
T. brucei	TbAQP1-3	x	x	x	dihydroxyacetone
T. cruzi	TcAQP	x	-	nd	nd
P. falciparum	PfAQP	x	x	x	sugar alcohols up to five carbons
T. gondii	TgAQP	x	x	x	

References: *T. brucei* (Uzcategui *et al.*, 2004), *T. cruzi* (Montalvetti *et al.*, 2004), *P. falciparum* (Hansen *et al.*, 2002), *T. gondii* (Pavlovic-Djuranovic *et al.*, 2003)

4.1.6 Aims of this study

This chapter aimed to characterise aquaporin genes in *P. nicotianae* (*PnAq*). Southern blot analysis was employed to determine the number of gene copies. In order to ascertain the expression profile of the *PnAq* genes RNA blot analysis of total RNA obtained from the four asexual developmental life cycle stages of *P. nicotianae* and a modified SSCP approach were carried out. For localisation studies, antibodies were raised against synthetic peptides and polyclonal antibodies were used in immunofluorescence assays.

4.2 Results

4.2.1 Sequencing and alignment of cDNA clone WS86

The cDNA clone WS86 was isolated during the differential screening of randomly selected clones from a 4h germinated cyst cDNA library of *P. nicotianae* by Dr Weixing Shan (Shan *et al.*, 2004). Plasmid DNA was isolated according to protocols described in Section 2.2.1 and was sequenced with pBS primers T3 and T7 from the 5' end and 3' end, respectively.

The cDNA clone WS86 was identified as a gene encoding an aquaporin after BLASTN and BLASTX analyses against sequences in GenBank. The 673 bp insert of WS86 encodes a polypeptide of 198 aa. A BLASTP search with the predicted protein sequence produced evidence of a conserved domain shared with a MIP superfamily in the NCBI database (Marchler-Bauer *et al.*, 2005, CDD Access Number 29423). This family includes several GLPs, a number of mammalian aquaporins (e.g. AQP1, 2 and 10) as well as TIPs, NIPs and PIPs of a range of plants. Overall, 76.8% of the WS86 amino acid sequence aligned with the conserved domain (Figure 4.9 and Figure 4.10). The WS86 amino acid residues from positions 3-185 align with the MIP conserved domain amino acid residues from positions 54-228. An X-ray diffraction structure of the *E. coli* glycerol facilitator Chain A representing the MIP superfamily proteins deposited in the PDB (1FX8, Fu *et al.*, 2000) was available through the CDD website. The program Cn3D (NCBI) was used to generate a 3D structure model of the aquaporin protein using the PDB file 1Fx8 as template (Figure 4.11).

In BLASTX GenBank database analysis, the hit with the highest homology to WS86 was an aquaporin 7-like protein of *Gallus gallus* (GenBank accession number XP_424498) with an *E* value of 6×10^{-43} (Figure 4.12). Significant hits were also observed for mammalian Aquaporins 3, 9 and 10 (e.g. human AQP10-AAH74896 and murine AQP3-AAH27400). The length of these homologous proteins is approximately 300 aa, suggesting that clone WS86 does not contain the full length cDNA for the protein in *P. nicotianae*. The protein encoded by cDNA clone WS86 was named PnAq for *Phytophthora nicotianae* Aquaporin.

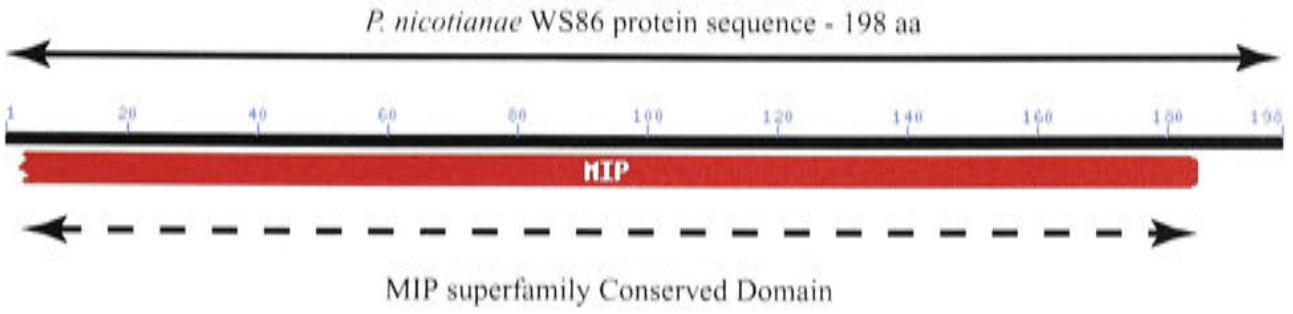


Figure 4.9 Conserved domain of the MIP family obtained by BLASTP search using the inferred amino acid sequence of cDNA clone WS86.

```
            CD-Length = 228 residues, only  76.8% aligned
            Score =  146 bits (369), Expect = 2e-36


Query:   3     HEGGANLNTAVTLAHCVYGRLPWWKAPGYMISQLLGAFCGAFIIYVMQYQNLNVIDPHRE   62
Sbjct:   54    HISGGHINPAVTLALAVGGRFPLIRVIPYIIAQLLGAILGAALLYGLYYGLYLEFLGANN   113

Query:   63    TTQSSFSTYPSDNISNYTAFYTEFIGTAMLVLSIYAITDKRNRSAGPVGSPFAFCLMIMA   122
Sbjct:   114   IVAGIFGTYPSPGVSNGNAFFVEFIGTFILVLVVFATTDDPNGPPPGGLAPLAIGLLVAA   173

Query:   123   LGMAFGMNTGYAVNPARDLGPRIFTAIAGWGSKVFTIRNYYFWIPLVADSLGGVCGAGLY   182
Sbjct:   174   IGLAGGPITGASMNPARSLGPALFTGLAR--------HWHYFWVYWVGPLIGAIAGALVY   225

Query:   183   RVL   185
Sbjct:   226   DYV   228
```

Figure 4.10 Alignment of the amino acid sequence derived from cDNA clone WS86 with the MIP conserved domain from the NCBI database. The WS86 amino acid residues from positions 3-185 aligned with the MIP conserved domain amino acid residues from 54-228 and produced an overall alignment of ~77% between the two sequences. The residues in red are identical in the two sequences whereas those in blue represent substitutions with conserved residues.

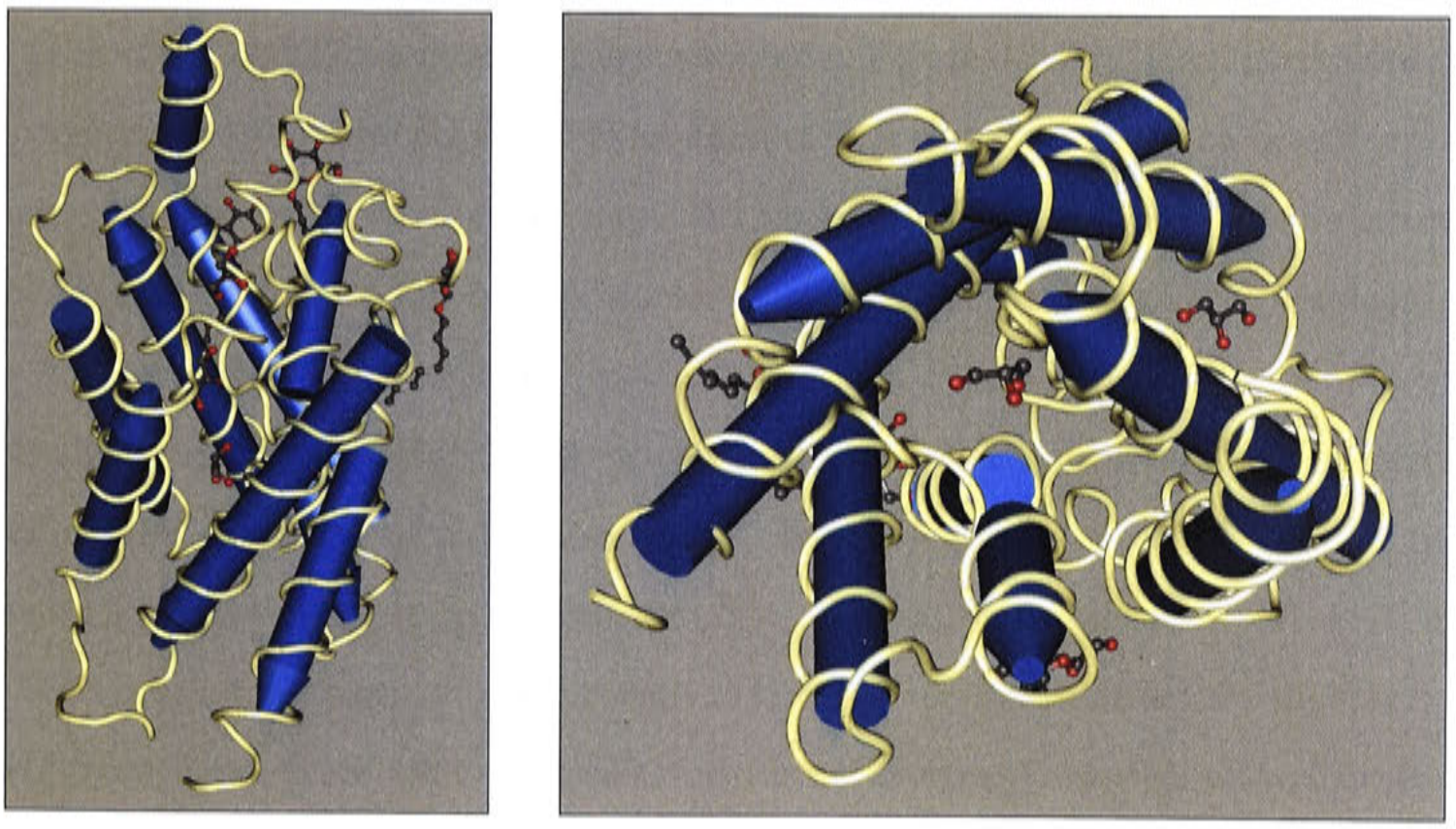


Figure 4.11 Structure of an aquaporin protein generated using the *E. coli* Chain A glycerol facilitator protein as a template from PDB / CDD (1FX8), displayed using the Cn3D viewer.

```
gi|50806738|ref|XP_424498.1|  PREDICTED: similar to Aquaporin 7
(Aquaporin-7 like) (Aquaporin adipose) (AQPap) [Gallus gallus]

Length=304


Score =  176 bits (447),  Expect = 6e-43
Identities = 87/193 (45%), Positives = 122/193 (63%), Gaps = 10/193 (5%)
Frame = +3

Query  18   GANLNTAVTLAHCVYGRLPWWKAPGYMISQLLGAFCGAFIIYVMQYQ--------NLNVI  173
            GA+LN A+TL HC++G LPW K P Y++ Q LG+F  A  I+ + Y           N  V
Sbjct  94   GAHLNAAITLTHCMFGNLPWRKLPVYLLGQFLGSFLAAATIFGLYYDALHDYTNGNFTVT  153

Query  174  DPHRETTQSSFSTYPSDNISNYTAFYTEFIGTAMLVLSIYAITDKRNRSAGPVGSPFAFC  353
             P    T S FSTYPS  +S    F+TEFI T ML+L I  I D++N  A
Sbjct  154  GP--TATASIFSTYPSPYMSLTGGFFTEFIATVMLLLGILVIHDEKNNGALKGTHALLTG  211

Query  354  LMIMALGMAFGMNTGYAVNPARDLGPRIFTAIAGWGSKVFTIRNYYFWIPLVADSLGGVC  533
            ++++ +GM  GMNTGYA+NP+RDL PRIFTAIAGWG  VF   NY++W+P+VA +LG +
Sbjct  212  ILVLGIGMGMGMNTGYAINPSRDLPPRIFTAIAGWGVDVFRAGNYWWWVPVVAPTLGSLA  271

Query  534  GAGLYRVLVEMHH  572
            G  ++++L+++H+
Sbjct  272  GGLIHKLLIDIHN  284
```

Figure 4.12 Alignment of the *P. nicotianae* cDNA clone, WS86, with the GenBank sequence showing the highest homology. The 673 bp insert of the cDNA clone shows high homology to a *G. gallus* aquaporin 7-like protein.

4.2.2 Isolation, subcloning and sequencing of a *PnAq* genomic clone

The insert of cDNA clone WS86 was used to screen a genomic *P. nicotianae* BAC library as described in Section 2.2.8. The first three lanes of the blot, representing genomic DNA digested with restriction enzymes *Bam*HI, *Eco*RI and *Hin*dIII, indicate that *PnAq* represents a gene family with at least four gene copies in *P. nicotianae*. Six out of seven lanes of pooled BAC DNA contain the *PnAq* gene (Figure 4.13 a). Two filters containing positive clones were screened with ^{32}P-labelled cDNA clone WS86. Both filters produced a single putative positive clone (Figure 4.13 b; 6A8 and 19I9. BAC clone screening experiments were carried out by Val Maclean). Restriction digests of DNA isolated from BACs 6A8 and 19I9 with enzymes *Eco*RI, *Hin*dIII and *Bam*HI and subsequent Southern blot analysis using the WS86 probe showed different restriction patterns in the EtBr stained gel but identical hybridisation results (Figure 4.14). As the hybridisation pattern was also identical to the genomic Southern blot analysis, it is likely that both BAC clones span the complete aquaporin gene family. The four bands obtained by digestion with *Bam*HI from BAC clone 19I9 were selected for subcloning into pBS vector. Only three fragments were successfully subcloned and sequenced in both directions after confirmation of the correct inserts by Southern blot analysis (Figure 4.15).

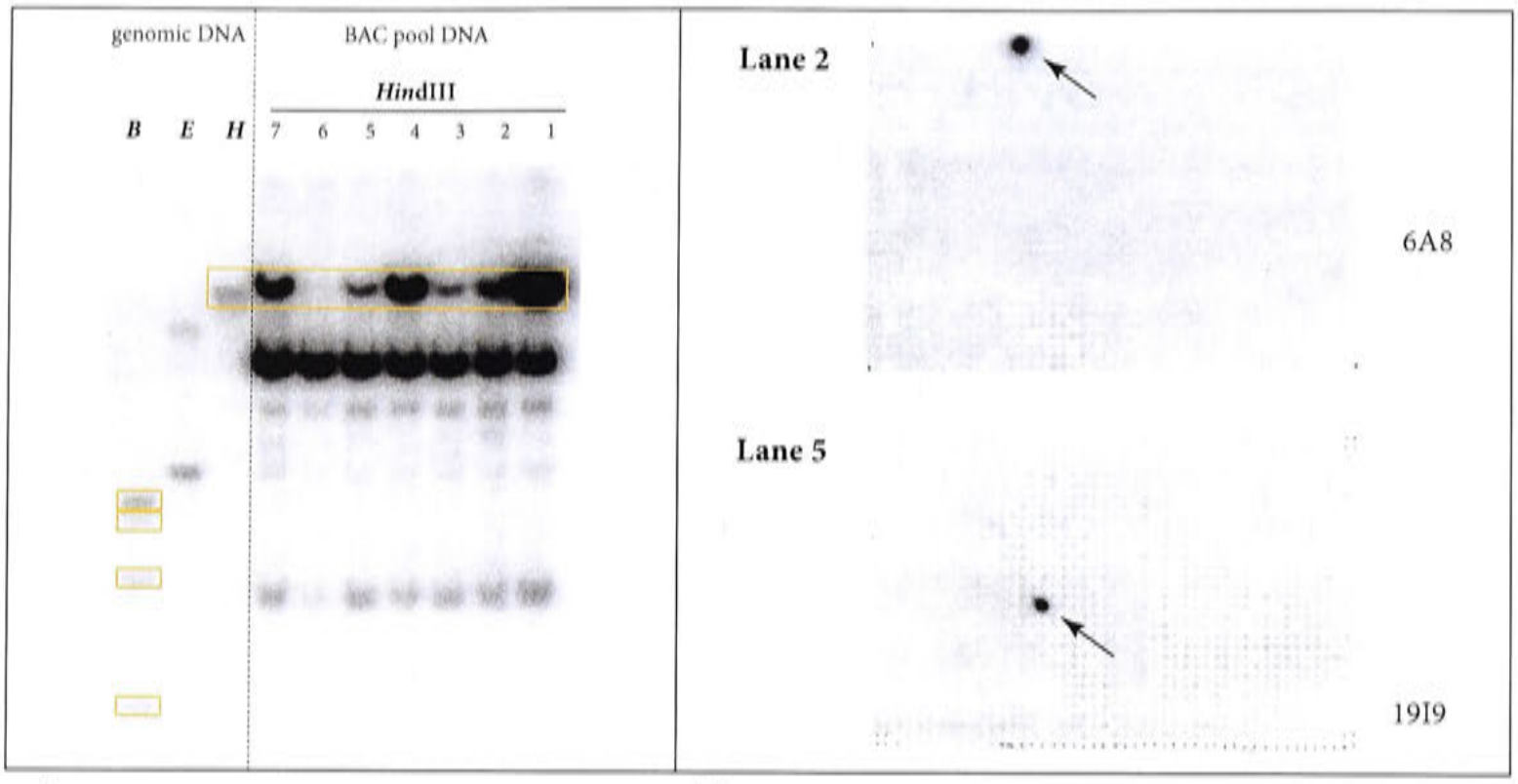


Figure 4.13 (a) Screening of pooled BAC DNA from a *P. nicotianae* genomic library digested with *Hin*dIII (lanes 1-7: each representing 1536 clones) and genomic DNA of *P. nicotianae* digested with restriction enzymes *Bam*HI (*B*), *Eco*RI (*E*) and *Hin*dIII (*H*) and probed with ^{32}P-labeled cDNA clone WS86. The signal in the *Hin*dIII digest lane in the pooled BAC DNA suggests that a genomic clone corresponding to WS86 is present in all BAC pools except 6 (upper orange box). The hybridisation result suggests that *PnAq* is a gene family in *P. nicotianae* consisting of at least four genes as indicated by the *Bam*HI digest (orange boxes). (b) Two membranes from the BAC pool (lanes 2 and 5) were hybridised with ^{32}P-labelled cDNA cloned WS86. Both filters produced a single putative positive clone (arrows, 6A8 and 19I9).

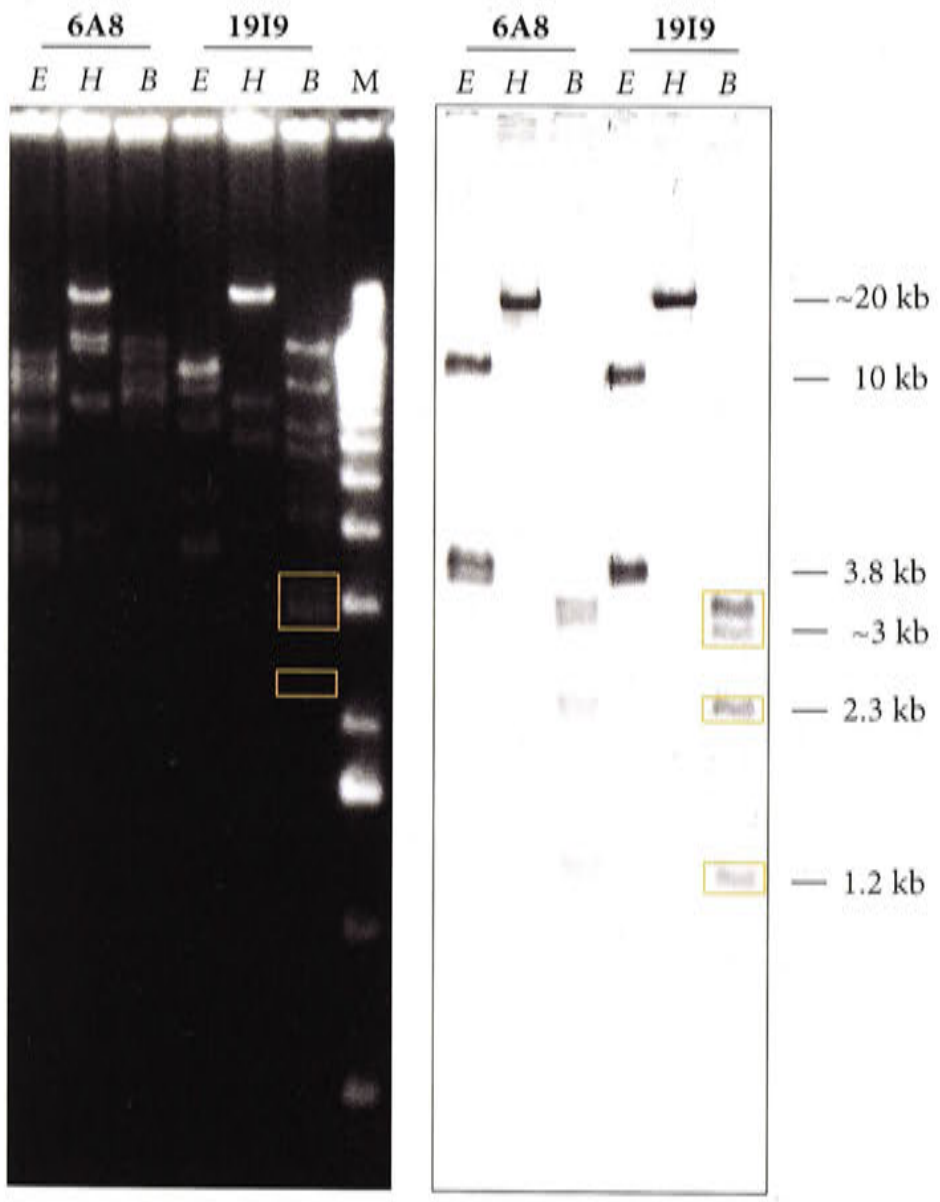


Figure 4.14 Restriction digests of two putative *PnAq P. nicotianae* BAC genomic clones with enzymes *Eco*RI (*E*), *Hin*dIII (*H*) and *Bam*HI (*B*) in order to identify restriction fragments for cloning the genomic clone into pBS vector. EtBr staining (left) of the gel shows a different digestion pattern in the two BAC clones, however hybridisation with cDNA probe WS86 shows an identical labelling pattern (right), indicating that both BAC clones span the whole gene cluster. The four *Bam*HI bands from clone 19I9 ranging from ~1.2 kb - ~3.4 kb (boxes in 19I9 *Bam*HI digest lane) were selected for subcloning in pBS to enable sequencing of the clones.

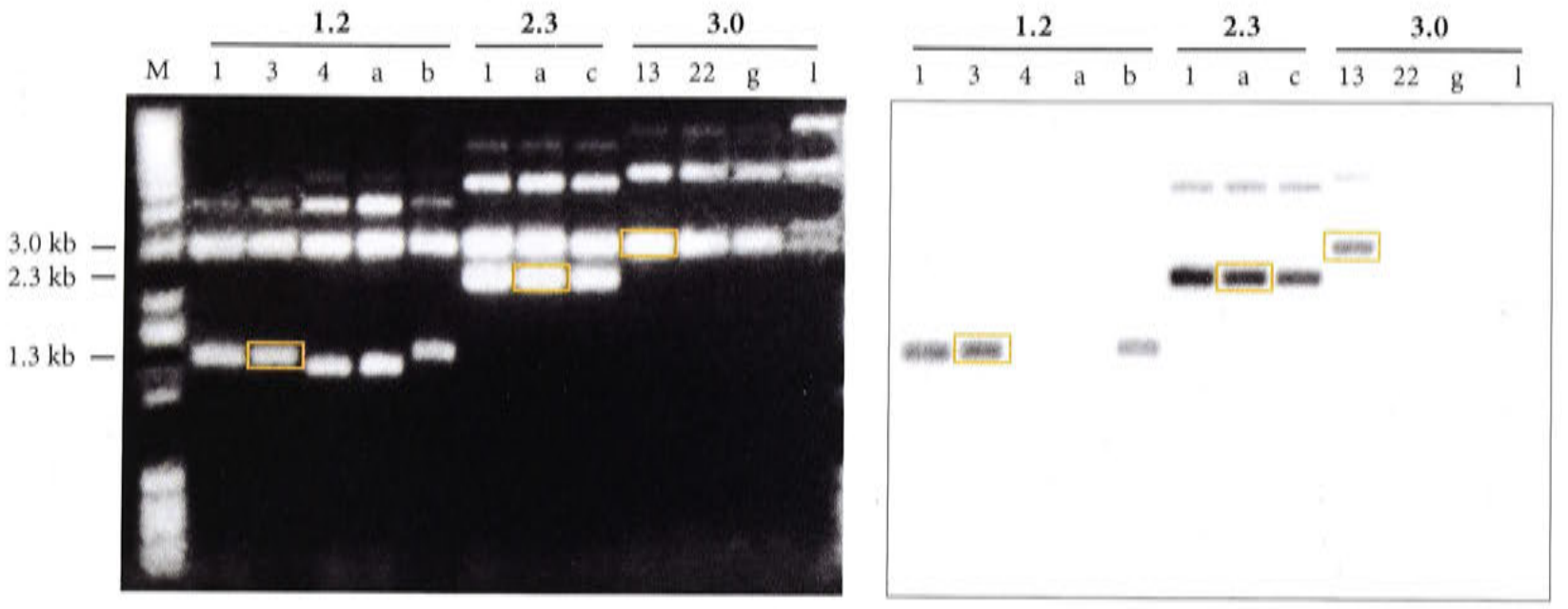


Figure 4.15 *Bam*HI restriction digests of several *PnAq* genomic subclones obtained from pBS + *PnAq* 19I9 *Bam*HI ligation reactions. The *Bam*HI digests of subclones representing the 1.2 kb, 2.3 kb and 3.0 kb fragments were probed with the ^{32}P-labelled cDNA clone WS86. Orange boxes indicate subclones that were sequenced. The picture on the left shows the EtBr stained gel, the picture on the right depicts the corresponding hybridisation results.

As subcloning of the fourth *Bam*HI fragment was not achieved, Southern blot analysis of BAC clone 19I9 was repeated with a different set of enzymes to identify further fragments suitable for subcloning. DNA was digested with *Eco*RI, *Sac*I, *Sal*I, *Xba*I and *Xho*I and the membrane hybridised with ^{32}P-labelled cDNA clone WS86 (Figure 4.16). The ~5 kb and ~6 kb fragments of the *Xho*I digest were selected for subcloning. Only one fragment was successfully subcloned into vector pBS and sequenced. Alignment of the sequences obtained showed that the *Xho*I fragment represented the same region as the three *Bam*HI fragments, meaning that a significant part of the genomic clone was still not encompassed by the subclones. Parts of the missing information were obtained by following a shotgun-like cloning approach. BAC clone 19I9 was digested with *Sal*I and the region of interest (~1 to 5 kb) was excised from the gel and ligated into a pBS vector. One hundred and twenty five colonies were subjected to colony PCR using pBS primers T3 and T7 and subsequent Southern blot analysis was performed with probe WS86 in order to identify insert-containing clones (data not shown). DNA from positive clones representing bands from the *Sal*I digested BAC clone 19I9 was isolated and fragments were fully sequenced. One pBS vector contained an insert that had two SalI fragments (pBS/S1_S2), as was shown by restriction enzyme analysis later (data not shown). Figure 4.17 shows the physical organisation of the *PnAq* gene family, the relative positions of the subcloned fragments, key restriction sites, as well as the predicted positions of the two fragments not subcloned (*Ba* and *Xa*). The organisation of subclone S2 could not be

fully resolved during the course of this thesis due to time constraints. It is conceivable, however, that the clone is located upstream of clone S4 considering approximate sizes of the fragments *Xa* and *Ba* of the *Xho*I and *Bam*HI digests, respectively. The small 425 bp *Sal*I fragment situated upstream of clone S4 was obtained by PCR amplification and sequencing of the PCR product. *P. nicotianae* aquaporin genes were named *PnAq1* to *PnAq5*. Alignment of cDNA clone WS86 with the ~9 kb genomic DNA contig showed that the clone corresponds to gene *PnAq2* and does not contain any introns.

The regions upstream of the *PnAq* genes were analysed for potential promoter binding sites. Figure 4.18 shows that all genes contain a variation of the 19 bp consensus motif GCYCATTYYNCAWTTTNYY (Pieterse *et al.*, 1994; McLeod *et al.*, 2004), commonly found in the promoter region of many oomycete genes. Furthermore, CT-rich regions, TATA-boxes and CAAT sequences were observed in some of the genes. One gene (*PnAq1*) also displayed a so called 'cold-box' motif, which was identified as a possible binding site for a zoosporogenesis-induced transcription factor (Tani and Judelson, 2006). Putative polyadenylation signals (consensus AATAAA), located downstream of the ORFs in the 3' UTR, were observed in all five genes.

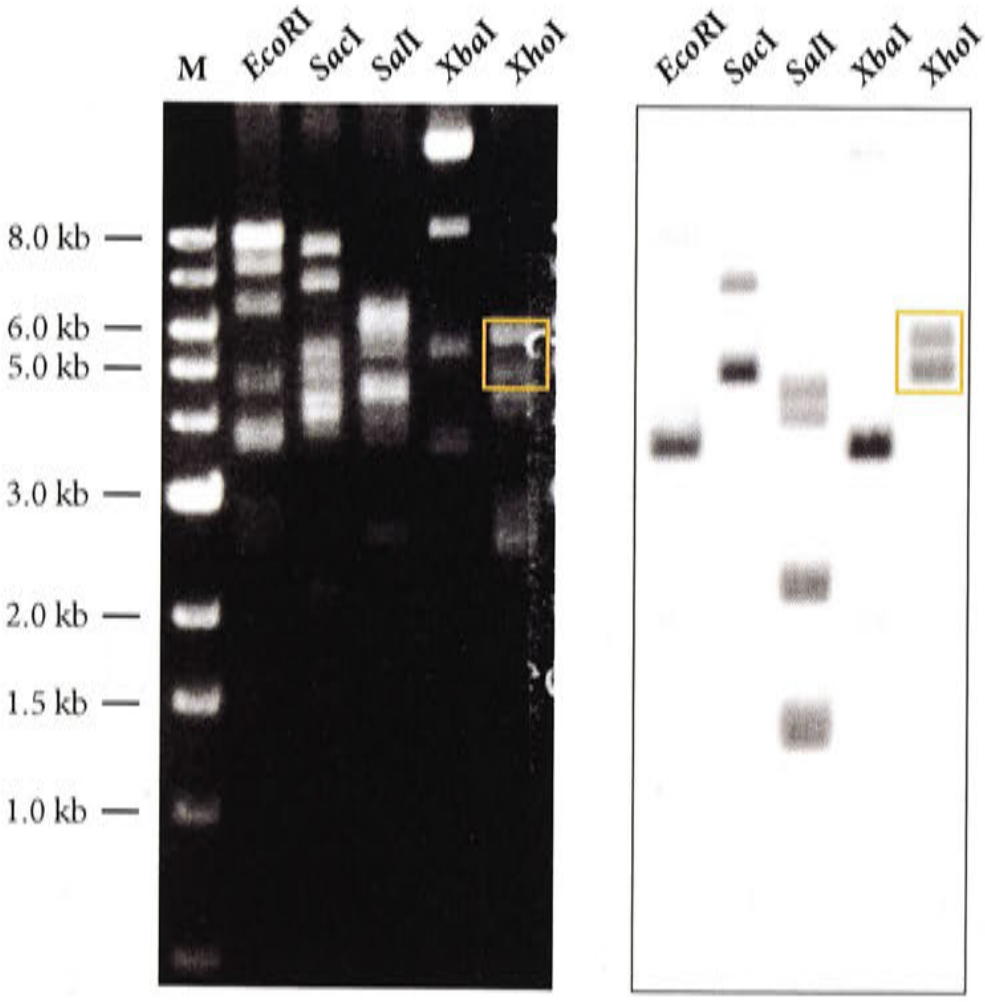


Figure 4.16 Southern blot analysis of *P. nicotianae* genomic clone 19I9 after digestion with five restriction enzymes, hybridised with cDNA clone WS86 in order to identify further restriction fragments for cloning into pBS vector. The ~5 kb and ~6 kb bands (orange box) from the *Xho*I digest were selected for subcloning.

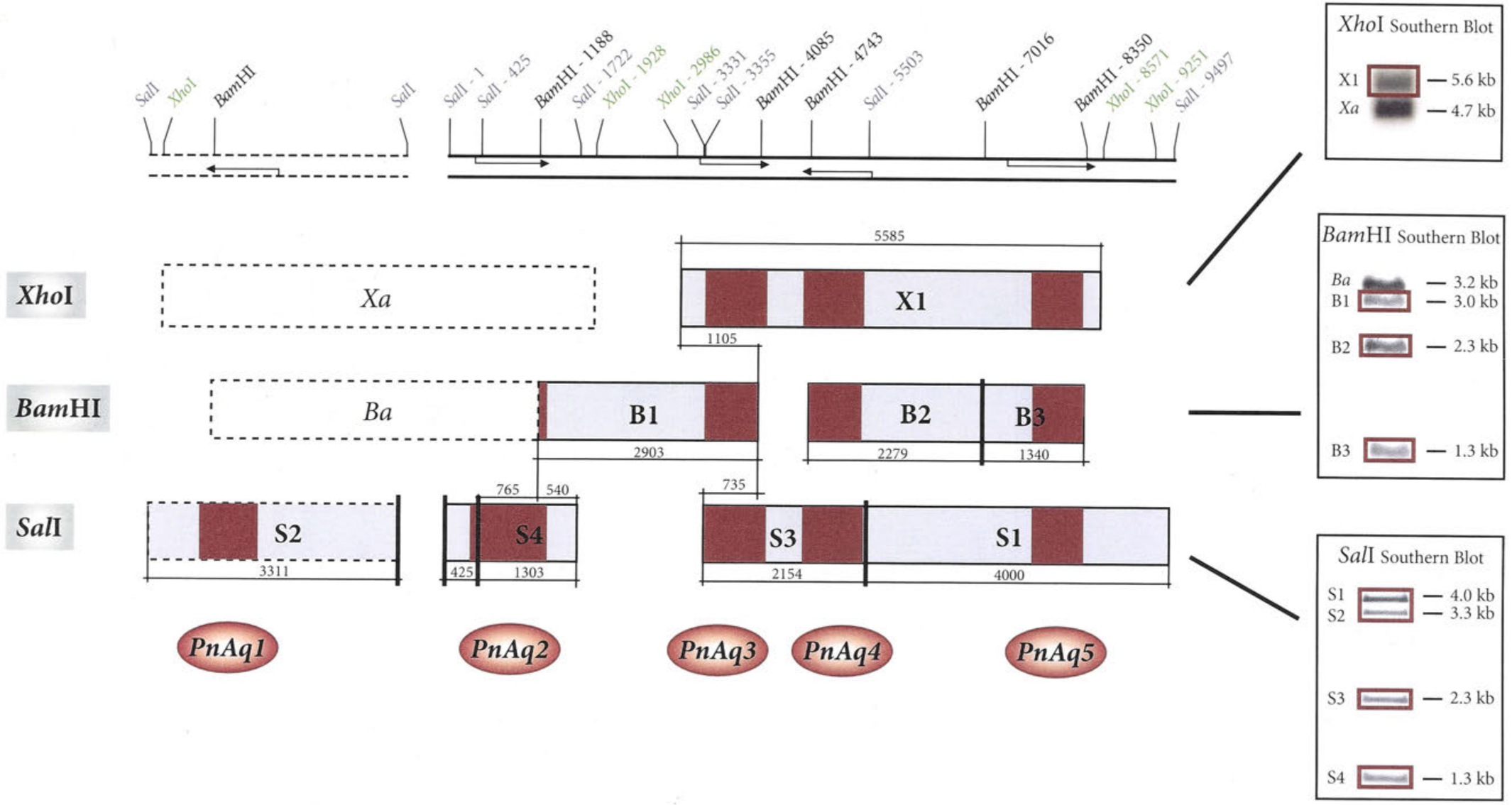


Figure 4.17 Diagram showing the organisation of the *PnAq* gene family determined by sequencing of several subclones from *P. nicotianae* BAC clone 19I9. Relevant restriction sites utilised for subcloning of fragments are shown on top. Southern blots on the right show the successfully subcloned and sequenced fragments with designated names (red boxes). Red solid boxes in the schematic presentation of the genomic contig depict aquaporin genes. The two boxes labelled *Xa* and *Ba* show the predicted positions of the fragments of *Bam*HI and *Xho*I digests which were not subcloned successfully. Numbers on the lines representing the genomic contig indicate length of fragments in bp.

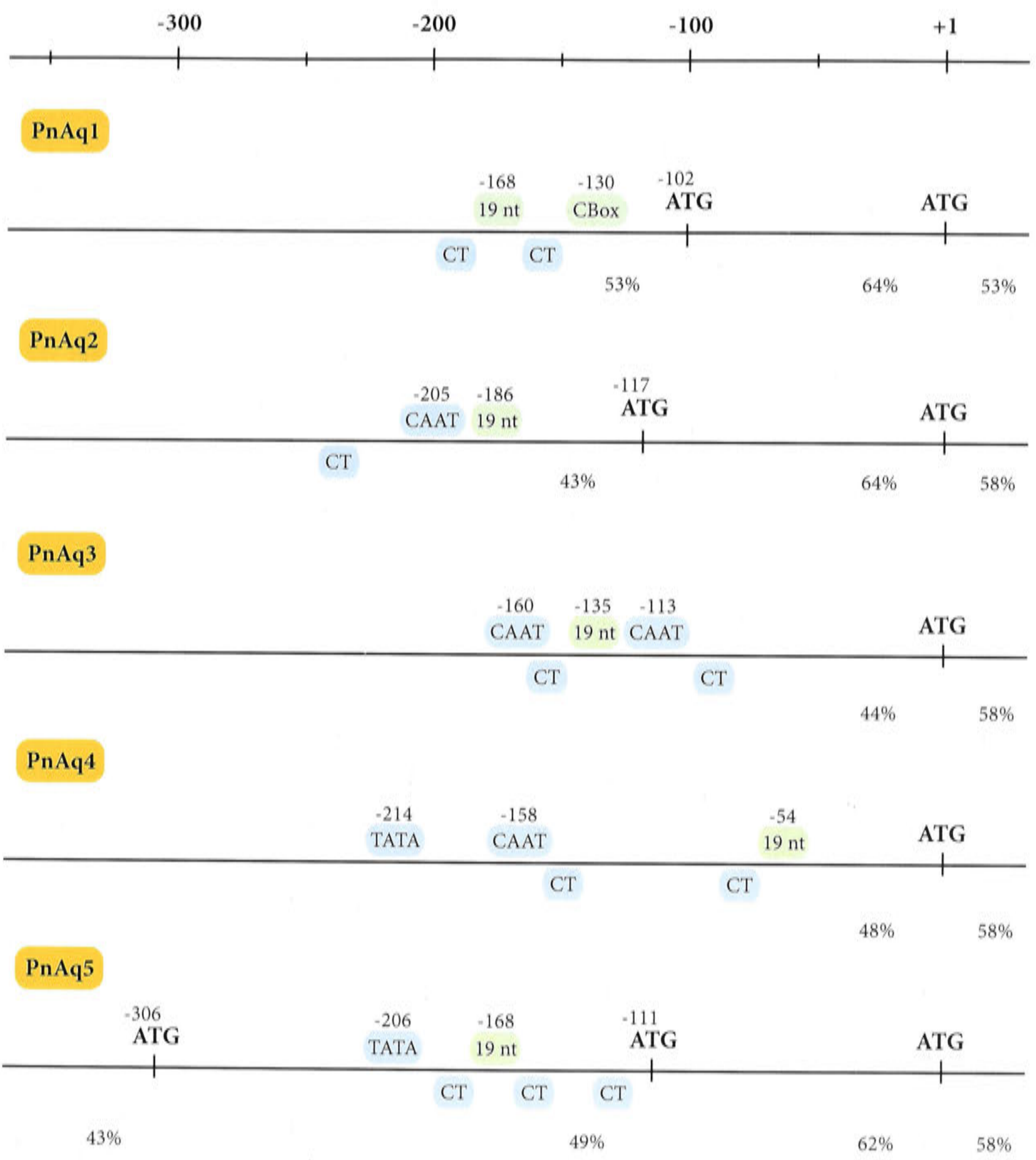


Figure 4.18 Promoter analysis of *PnAq* genes. Potential promoter binding sites are displayed for each *P. nicotianae* aquaporin gene. Numbers on top denote nucleotides. Percentages below each gene indicate G+C content of regions. 19nt–*Phytophthora* specific promoter motif, CT–CT rich region, CBox–"Cold-box" motif of *Phytophthora*.

4.2.3 PnAq proteins, homologies and conserved domains

Five ORFs were identified in the genomic clones analysed (Figure 4.17 and Figure 4.18, see Appendix G for sequences). Genes *PnAq3* and *PnAq4* encode proteins of 298 aa each. It was difficult to determine the site of the correct translation start codon for genes *PnAq1* and *PnAq5* due to the absence of cDNA clones for those genes. Two potential ATG start codons positioned 102 bp apart were observed in the correct ORF for *PnAq1*, encoding for either a 277 or 311 aa protein. *PnAq5* shows three potential translation start sites, encoding either a 377, 312 or 275 aa protein. Based on promoter analysis of the genes it is likely that the first ATG of *PnAq1* and the central ATG of *PnAq5* (-102 and -111, respectively, in Figure 4.18) serve as translation start codons. Gene *PnAq2* also shows two in-frame start codons. As the cDNA clone corresponding to *PnAq2*, WS86, does not carry the full-length cDNA information, the correct translational start site was also inferred by promoter analysis. It is likely, considering distances between promoter binding sites and potential ATGs and G+C contents of regions, that the first ATG (-117) serves as start site.

Predicted molecular weights and pIs of the DNA sequences inferred PnAq proteins are shown in Table 4.7. Identity and similarity of protein sequences was evaluated using the program MatGAT (Matrix Global Alignment Tool, Campanella *et al.*, 2003). Proteins show 71-84% identity and 81-88% similarity to each other (Table 4.8). To further analyse the proteins, sequences were subjected to multiple sequence alignment analysis using ClustalX, and several transmembrane prediction programs and a hydropathy plot were used to predict the secondary structure (Table 4.9, Figure 4.19, Figure 4.20). All proteins show the typical aquaporin structure with six transmembrane helices (Figure 4.20).

Table 4.7 Prediction of molecular weights and pIs of *P. nicotianae* aquaporin proteins.

Protein	Length in aa	Predicted MW in kDa	Predicted pI
PnAq1	277 | **311**	30.2 | **33.9**	6.99 | **5.82**
PnAq2	278 | **317**	30.5 | **34.4**	8.14 | **6.18**
PnAq3	298	32.6	6.07
PnAq4	298	32.6	6.06
PnAq5	275 | **312** | 377	30.1 | **33.9** | 41.1	6.94 | **5.29** | 5.78

Table 4.8 Global sequence identity/similarity between the five *P. nicotianae* aquaporin proteins. Percentages were obtained using the program MatGAT (Campanella *et al.*, 2003) with the following settings: Matrix–BLOSUM62, First gap–12, Extending gap-2.

		Identity in %				
		PnAq1	**PnAq2**	**PnAq3**	**PnAq4**	**PnAq5**
Similarity in %	**PnAq1**		75	71	74	75
	PnAq2	86		74	84	84
	PnAq3	81	83		76	74
	PnAq4	85	88	87		84
	PnAq5	84	87	82	88	

Table 4.9 Prediction of transmembrane regions in PnAq2 by five prediction programs. (TMpred - Hofmann and Stoffel, 1993, SOSUI - Hirokawa *et al.*, 1998, TMHMM - Krogh *et al.*, 2001, Hmmtop - Tusnady and Simon, 2001, TopPred - Claros and von Heijne, 1994)

	TMpred		**SOSUI**		**TMHMM**		**Hmmtop**		**TopPred**	
	from	**to**	**from**	**to**	**from**	**to**	**from**	**to**	**from**	**to**
1	71	90	71	93	66	88	71	88	70	90
2	103	119	103	125	98	120	101	118	100	120
3	149	170	148	170	148	170	149	168	150	170
4	200	220	198	218	197	219	199	218	200	220
5	226	248	227	249	226	248	231	250	229	249
6	281	305	283	305	283	305	281	298	281	301

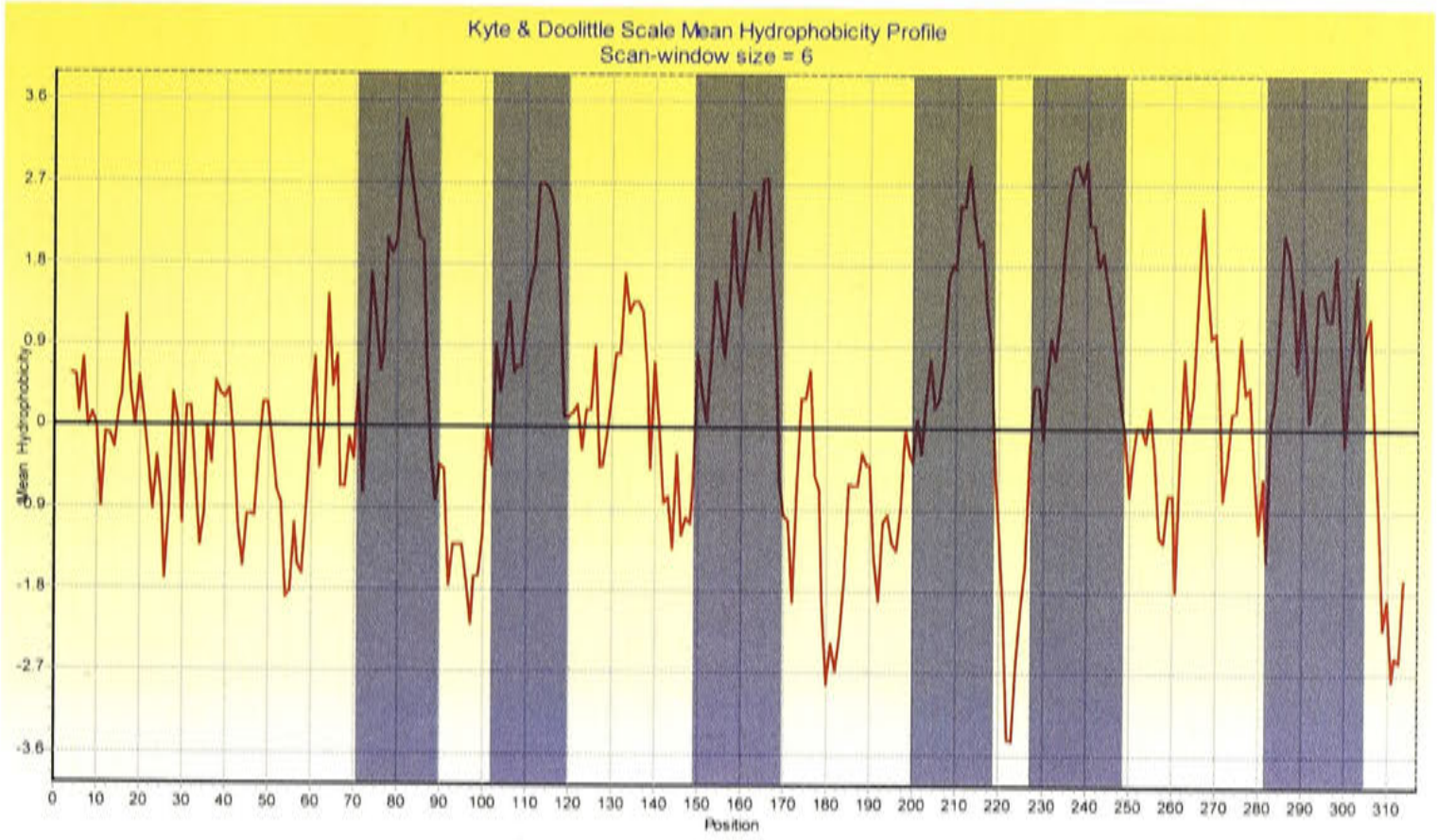


Figure 4.19 The Kyle and Doolittle hydropathy profile shows the hydrophobic and hydrophilic regions of the PnAq2 amino acid sequence (317 aa). The plotted regions below '0' (values on the Y-axis) constitute hydrophilic amino acids in the sequence while the plotted regions above '0' (values on the Y-axis) constitute hydrophobic amino acids in the sequence. Grey areas show predicted transmembrane regions as predicted by several programs (Table 4.9).

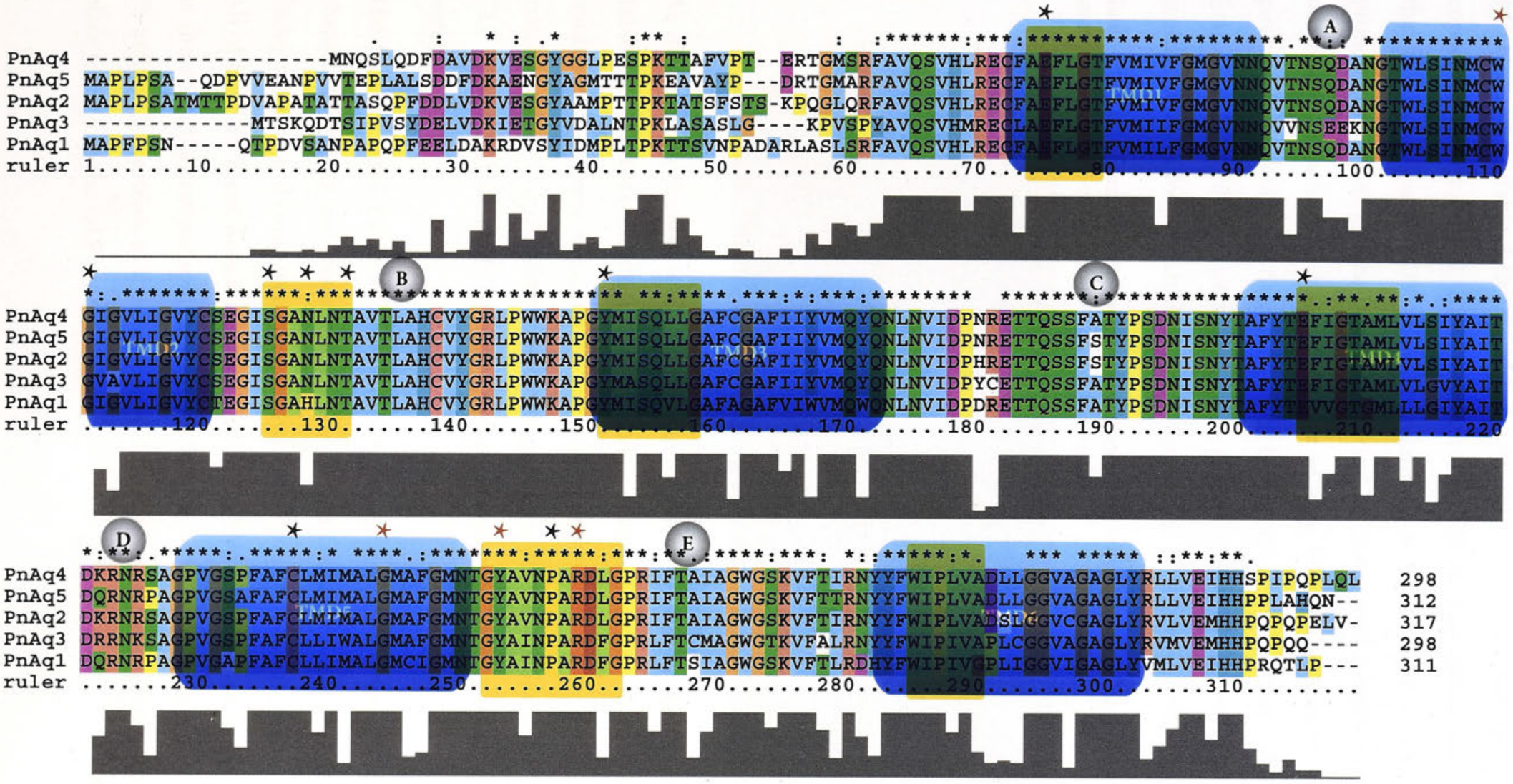


Figure 4.20 PnAq amino acid multiple sequence alignment with sequence motifs. Orange boxes indicate conserved domains. Red (big) asterisks denote amino acid residues involved in formation of the narrowest constriction, whereas black (big) asterisks indicate residues that play a role in stabilisation of the pore by side-chain interaction. The overlay shows predicted TMD and loops A-E. See text for details.

Furthermore, a number of conserved domains were observed in all PnAq proteins as summarised in Table 4.10 (see also Figure 4.20). There are no changes in any of the conserved motifs located within the first four TMDs. The conserved glycine in TMD5 that occurs in many MIPs in other organisms is replaced by a cysteine in all PnAq proteins. In TMD6, glycine is only conserved in PnAq1 and is replaced by alanine in the other four sequences. The first NPA box in loop B shows a substitution of the second position from proline to threonine in all five proteins. Furthermore, the histidine in the fourth position of the motif is changed to asparagine in all proteins but PnAq1. The motif located in loop E, which contains the second NPA box, is conserved in all proteins.

Table 4.10 Conserved domains and their corresponding sequences in PnAq proteins. Underlined italic amino acids indicate substitution of important residues.

Location	Conserved motif	Corresponding sequence in PnAq proteins
TMD1	AEFxxT	AEFLGT
TMD2	G	G
TMD3	YxxxQxxG	YM(I/A)SQ(L/V)LG
TMD4	ExxxTxxL	E(F/V)GT(A/G)ML
TMD5	G	C
TMD6	W…PxxG	WIP(L/I)V(*A*/G)
LoopB	SGxHxNPAVT	SGA(*N*/H)LN*T*AVT
LoopE	GxxxNPAR(S/D)xG	GYA(V/I)NPARD(L/F)G

As discussed in Section 4.1.2, the selectivity filter of MIPs is shaped by four residues. Table 4.11 lists these residues from a number of organisms and the corresponding PnAq residues which were obtained by sequence comparison. Three of the PnAq residues match the *E. coli* GlpF residues, suggesting that the proteins in *P. nicotianae* belong to the GLP cluster and are likely to transfer glycerol and possibly also other small solutes and water.

The five positions P1 to P5, proposed by Froger and co-workers as determinants for channel selectivity and classification into clusters (Froger *et al.*, 1998), were identified in PnAq proteins and compared to representatives from each cluster as shown in Table 4.12. All of these residues were conserved within the five *P. nicotianae* PnAq genes. However, as shown in the comparison, they do not follow the P1 to P5 rule as residues P1, P4 and P5 assign the proteins to the aquaporin cluster, whereas residues at positions P2 and P3 place them into the GLP group.

Table 4.11 Residues forming the selectivity filter in different organisms as inferred from available X-ray 3D structure. Corresponding residues of *P. nicotianae* aquaporins were obtained by protein sequence comparison. Denoted residue numbers correspond to protein PnAq2. ∅-narrowest part of the constriction

MIP	Organism	Reference	Position ①	②	③	④	∅
PnAq	***P. nicotianae***	-	**W109**	**G243**	**R258**	**Y252**	-
AQP1	Human	Murata *et al.*, 2000	F56	H180	R195	C189	3 Å
	Bovine	Sui *et al.*, 2001	F58	H182	R197	C191	2.8 Å
AQP0	Sheep	Gonen *et al.*, 2004	F48	H172	R187	A181	~2 Å
AQPZ	*E. coli*	Savage *et al.*, 2003	F43	H174	R189	T183	~2 Å
AqpM	*M. marburgensis*	Lee *et al.*, 2005	F62	I187	R202	S196	2.5 Å
GlpF	*E. coli*	Fu *et al.*, 2000	W48	G191	R206	F200	3.8 Å

Table 4.12 Differentiation of the two distinct clusters (Aquaporins and GLPs) based on the five positions P1-P5 as proposed by Froger *et al.*, 1998. AQP1 (human) and GlpF (*E. coli*) residues are shown as examples for residues considered to be specific for the respective cluster. Residues from the *A. thaliana* NIP family (Quigley *et al.*, 2001) represent plant NIPs which commonly do not follow the P1-P5 rule. Denoted PnAq residues are conserved within the *P. nicotianae* aquaporin group.

Pos.	Location	Cluster	Discrimination	AQP1	GlpF	NIPs	PnAq
P1	TMD3/ loop C	AQP	non-aromatic	T			Q
		GLP	aromatic		Y	F/Y	
P2	loop E	AQP	small & uncharged	S		S/T	
		GLP	Asp (D)		D		D
P3	loop E	AQP	small & uncharged	A		A	
		GLP	Lys/Arg (R/K)		K		R
P4	TMD 6	AQP	aromatic	F		Y	F
		GLP	Pro (P)		P		
P5	TMD 6	AQP	aromatic	W			W
		GLP	non-aromatic		L	I/L/M	

The positions of the pore forming loops are stabilised through a number of ion pairs and hydrogen bonds of conserved residues. Table 4.13 shows a list of residue interactions as inferred from the structure of AQP1 and the corresponding PnAq residues which were obtained by sequence comparison.

Table 4.13 Residue interactions involved in the stabilisation of the pore in AQP1 and PnAq proteins, as inferred from the 3D structure of human AQP1 and sequence comparison, respectively. Denoted PnAq residue numbers correspond to protein PnAq2. (Source: Murata *et al.*, 2000)

	Residue			
	AQP1	**PnAq**	**Location**	**Type of interaction**
A	P77	**T130**	loop B (NPA box)	Van der Waals
	P193	P256	loop E (NPA box)	
B	H74	H127 (PnAq1) **N127 (PnAq2-5)**	loop B	Ion pair
	E17	E74	TMD 1	
C	R195	R258	loop B (helix)	Salt bridge
	E142	E204	TMD 4	
D	S71	S124	loop B	Hydrogen bond
	Y97	**T150**	TMD 3	
E	G57	G110	TMD 2	Van der Waals
	G173	**C236**	TMD 5	

4.2.4 Northern blot analysis of cDNA clone WS86

To characterise the expression profile of *PnAq* genes in *P. nicotianae*, RNA blot analysis was carried out. The RNA blot was prepared from total RNA of four life cycle stages of *P. nicotianae*: vegetative hyphae, 4 h sporulating hyphae, zoospores and 2 h germinated cysts. The ^{32}P-labelled WS86 probe revealed a transcript of ~ 1.4 kb in vegetative hyphae and germinated cysts. The *PnAq* transcript is upregulated in germinated cysts but not detected in sporulating hyphae and zoospores. Equal loading of RNA was demonstrated by methylene blue staining of rRNA and hybridisation with the constitutively expressed cDNA clone WS41 (Figure 4.21, experiments were carried out by Val Maclean and Weixing Shan; Shan *et al.*, 2004).

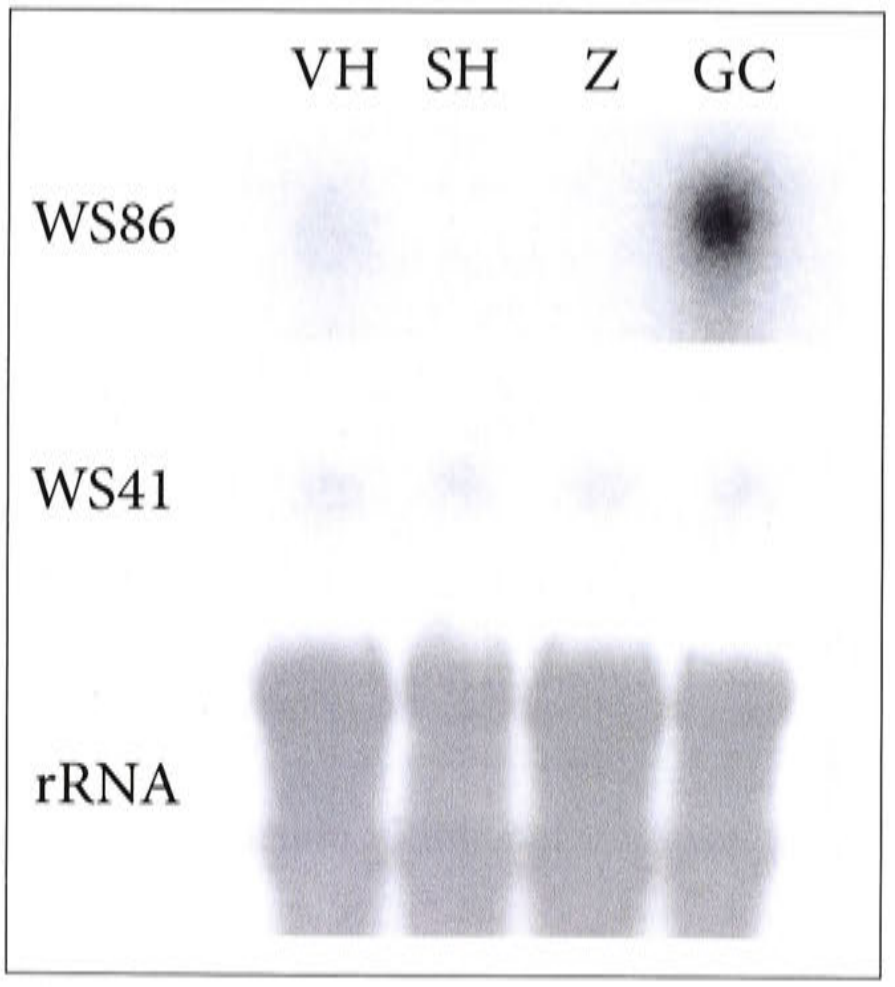


Figure 4.21 Northern blot analyses with RNA from four life cycle stages of *P. nicotianae* probed with ^{32}P-labelled cDNA clone WS86. The hybridisation result indicates *PnAq* expression in vegetative hyphae (VH) and germinated cysts (GC) but not in sporulating hyphae (SH) and zoospores (Z). Expression is highly up-regulated in germinated cysts. Staining of rRNAs with methylene blue (rRNA) and hybridisation with ^{32}P-labelled cDNA clone WS41, a housekeeping gene constitutively expressed in all four developmental stages, show quality and quantity of total RNA. (Source: Shan *et al.*, 2004)

4.2.5 *PnAq* expression study using a modified SSCP approach

The WS86 cDNA is derived from aquaporin gene *PnAq2* and was used as a probe in Southern and northern blot analysis. Comparison of the WS86 cDNA sequence with the corresponding DNA stretches of the other four genes revealed 77-93% similarity between the sequences (using MatGAT). As the WS86 cDNA cross-reacted with all the *PnAq* genes in the Southern blot analysis, it seemed likely that it would also cross-react with mRNA transcripts from all the *PnAq* genes in RNA blots. In northern blot experiments (see above), only two developmental stages (vegetative hyphae and germinated cysts) showed the presence of aquaporin mRNA transcripts. However, depending on conditions used in northern blot analysis (e.g. stringency for washes), low abundance copies which display low similarity to the probe might not be picked up. Furthermore, it cannot be inferred from the RNA blot which aquaporin genes are expressed in vegetative hyphae and germinated cysts. In order to address these issues, a modified SSCP experiment was carried out to analyse the expression profile of all five aquaporin genes in the four developmental stages of *P nicotianae*. SSCP is a method that has the potential to distinguish single base pair differences between DNA fragments. SSCP is based on the principle that the electrophoretic mobility of a single stranded DNA molecule in a non-denaturing gel is dependent on its size and conformation. The length of a strand, location and number of regions of base pairing determine the secondary and tertiary conformations of the molecule. These conformations are highly dependent on primary sequence. Hence, a mutation at a particular site in the primary sequence can change the conformation of the molecule, which alters its mobility during electrophoresis (Gasser and Monti, 1997). Generally, the region of interest in the genome or cDNA is amplified by PCR, denatured to form single strands, and analysed by non-denaturing polyacrylamide gel electrophoresis (Orita *et al.*, 1989). The procedure of the SSCP expression pattern analysis of aquaporins in *P. nicotianae* is shown in the flow-chart in Figure 4.22.

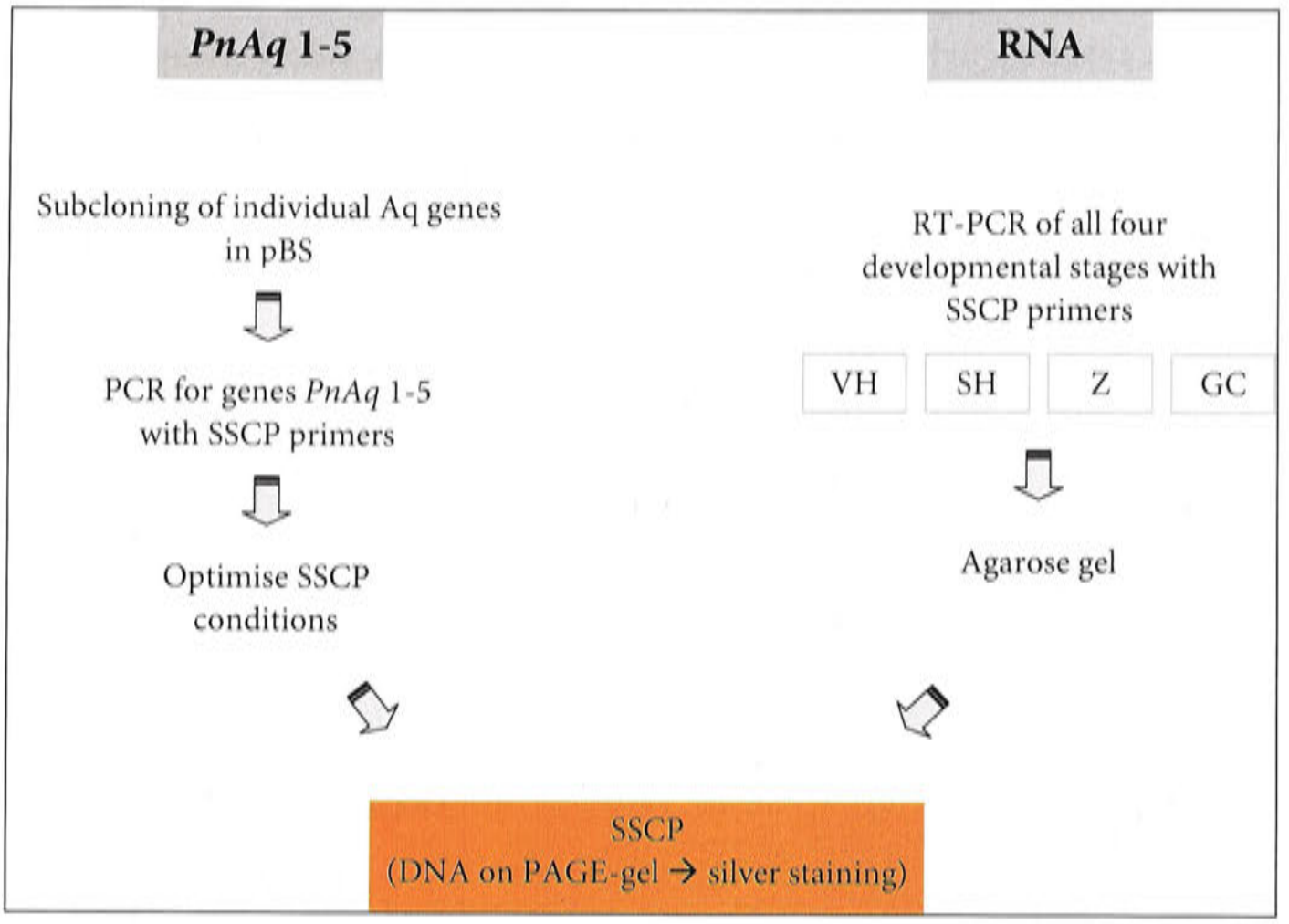


Figure 4.22 Flow-chart for the expression study of *P. nicotianae* aquaporin genes using a modified SSCP approach. For further details see text. VH-vegetative hyphae, SH-sporulating hyphae, Z-zoospores, GC-germinated cysts

All genes, except for *PnAq2* which is the sole gene in fragment S4, were individually subcloned into pBS vector using genomic subclones pBS_S1/S2 and pBS_X1. Table 4.14 and Figure 4.23 show the genes of interest and the restriction enzyme sites utilised. A multiple sequence alignment of all five aquaporin genes was generated and used for the design of a suitable primer pair, which allows for simultaneous amplification of a PCR product from all five genes. Two conserved regions producing a 274 bp PCR product with the primer pair FOR SSCP and REV SSCP were identified (Figure 4.24). SSCP analysis on PAGE-gels was optimised using PCR products from all five genes in a series of experiments to achieve a good resolution of the reference samples. Several factors which are thought to influence DNA separation on PAGE-gels, such as running time, temperature, buffer composition and gel concentration, were adjusted until sufficient resolution was observed.

RNA was prepared from four asexual lifecycle stages of *P. nicotianae* as described in Section 2.4.1 and a negative control experiment lacking the enzyme RT was carried out. No contaminating genomic DNA was observed (Figure 4.25). RNA was then subjected to reverse transcription using primer REV SSCP, followed by PCR amplification with primers FOR SSCP and REV SSCP. This allowed amplification of PCR products of all

expressed aquaporin genes simultaneously due to the fact that primers annealed to all five genes. DNA concentrations were determined after gel purification and approximately equal amounts of DNA were analysed on an agarose gel (Figure 4.26). DNA amounts were adjusted based on the gel result and an SSCP experiment was carried out. Figure 4.27 shows the running patterns of the purified PCR products of the five individual aquaporin genes (*PnAq1-5*) and the samples of the four developmental stages. Based on the bands present it was concluded that aquaporin genes *PnAq2* and *PnAq4* are expressed in all four developmental stages of *P. nicotianae*. Additionally, gene *PnAq3* is expressed in zoospores only. *PnAq1* expression was observed in vegetative and sporulating hyphae and to some extent in germinated cysts. There was no evidence of expression of *PnAq5* (Table 4.15).

Table 4.14 Subcloning of individual aquaporin genes into pBS vector. Genes were excised from the indicated clone (refer to Figure 4.17), digested with restriction enzymes (RE) shown and subcloned into pBS vector. *PnAq2* is present in fragment S4 as the sole gene and subcloning was therefore not required.

Gene	Excised from clone	RE	Size bp
PnAq1	S1_S2	*StuI*/*PmlII*	1408
PnAq 3	X1	*SpeI*	1165
PnAq 4	X1	*HinfI*	1099
PnAq 5	X1	*KpnI*/*XhoI*	1922

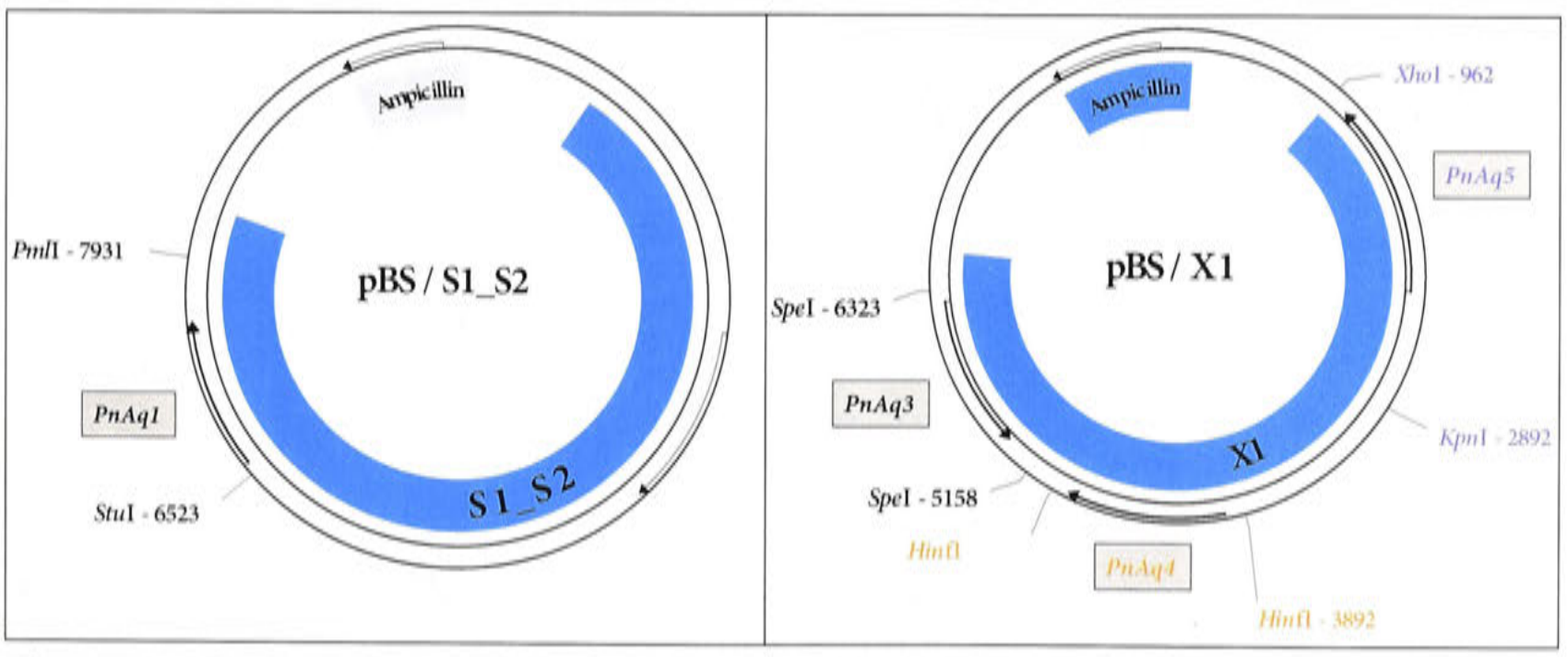


Figure 4.23 Schematic diagram of pBS vector containing *P. nicotianae* BAC clone 19I9 fragment S2 (a) and X1 (b) (refer to Figure 4.17). Shown are relevant restriction sites for the subcloning of individual genes into pBS vector as described in Table 4.14.

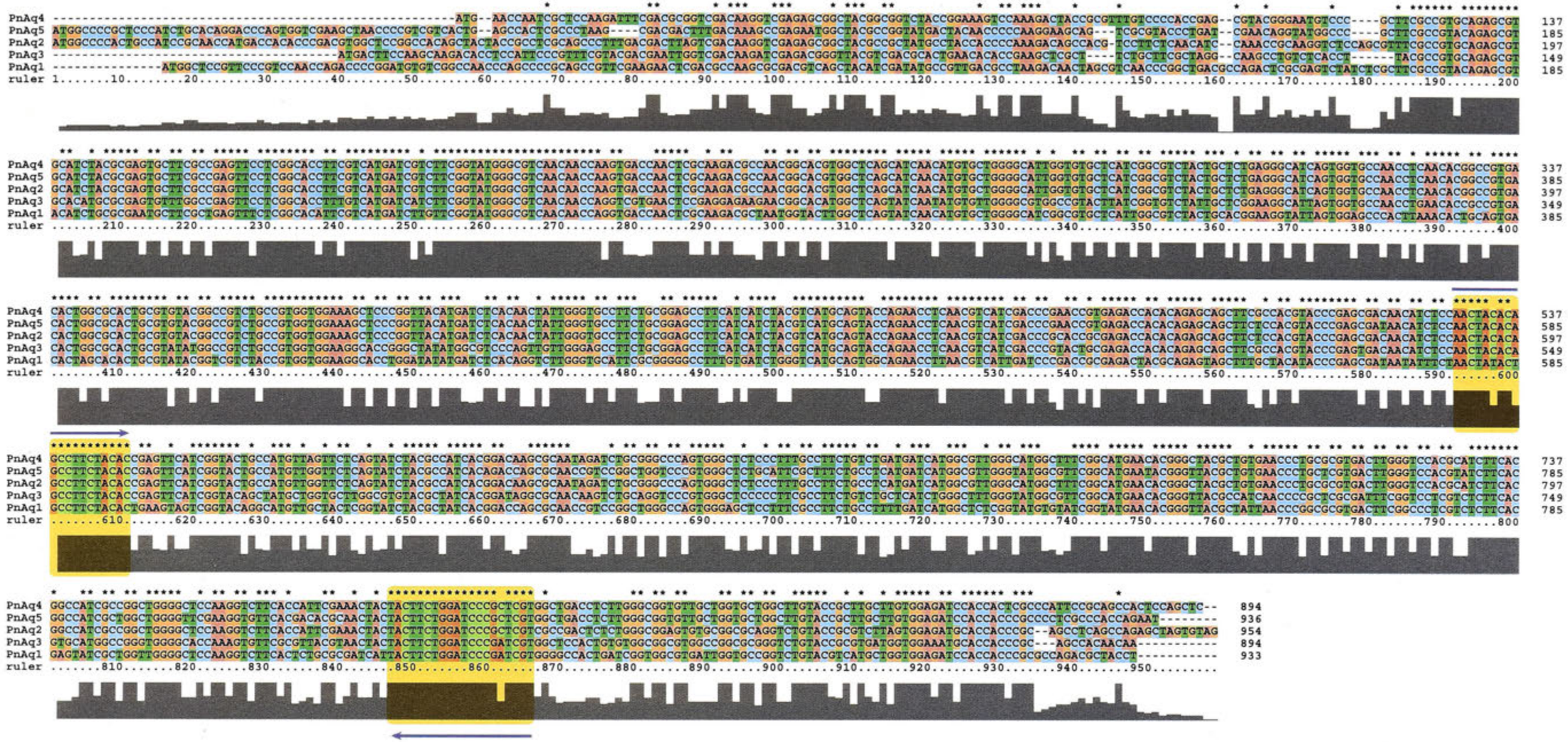


Figure 4.24 Multiple DNA sequence alignment of *P. nicotianae PnAq* genes. Highlighted regions with arrows represent primer positions (FOR SSCP and REV SSCP).

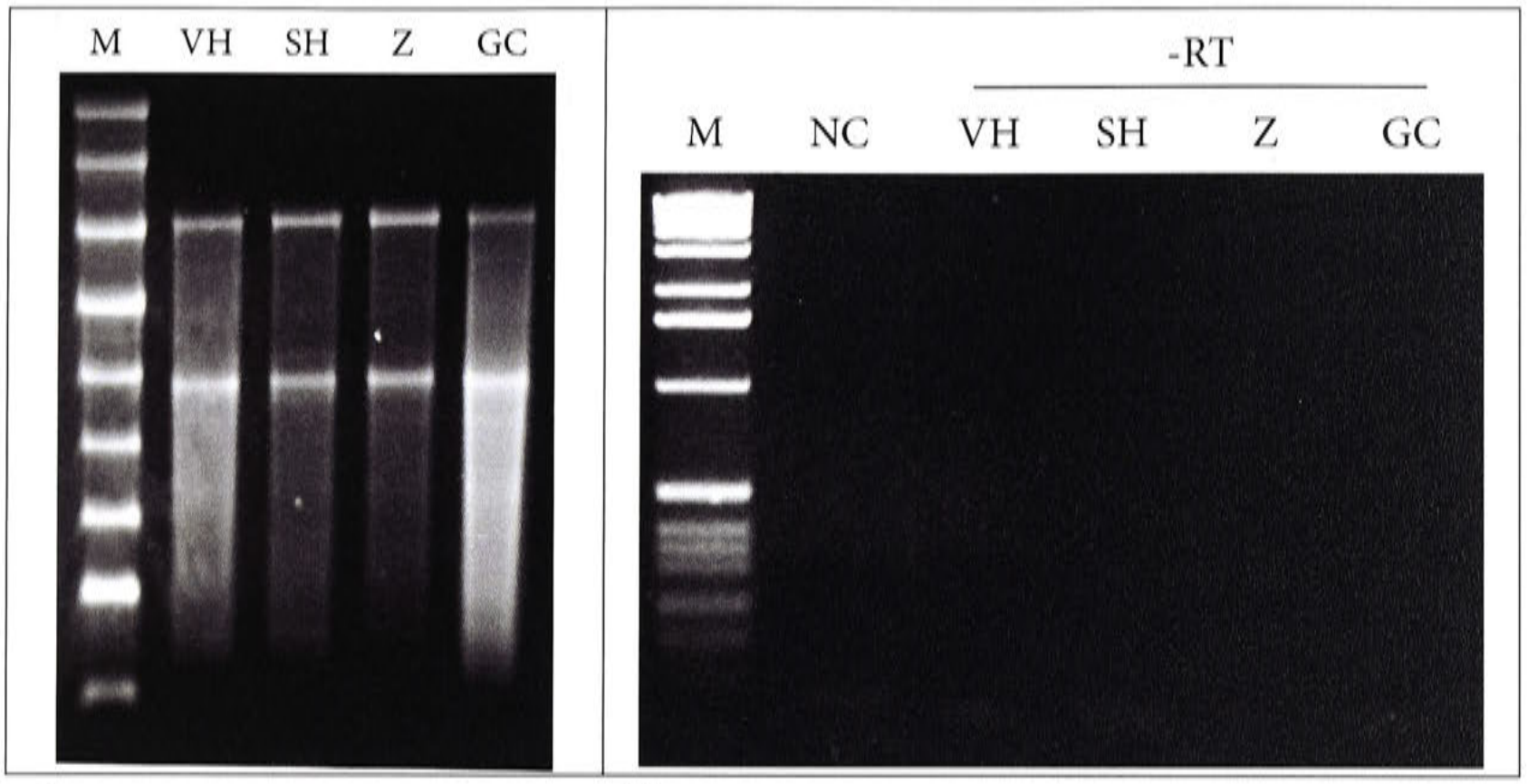


Figure 4.25 RNA isolation and reverse transcription negative control for SSCP analysis of *P. nicotianae PnAq* genes. (a) 10 µg total RNA of four developmental stages of *P. nicotianae*. (b) RNA was subjected to a control PCR using primers FOR SSCP and REV SSCP. No signal was observed, indicating that the RNA preparations were free of genomic DNA. M–Marker, NC–negative control, VH–vegetative hyphae, SH–sporulating hyphae, Z–zoospores, GC–germinated cysts

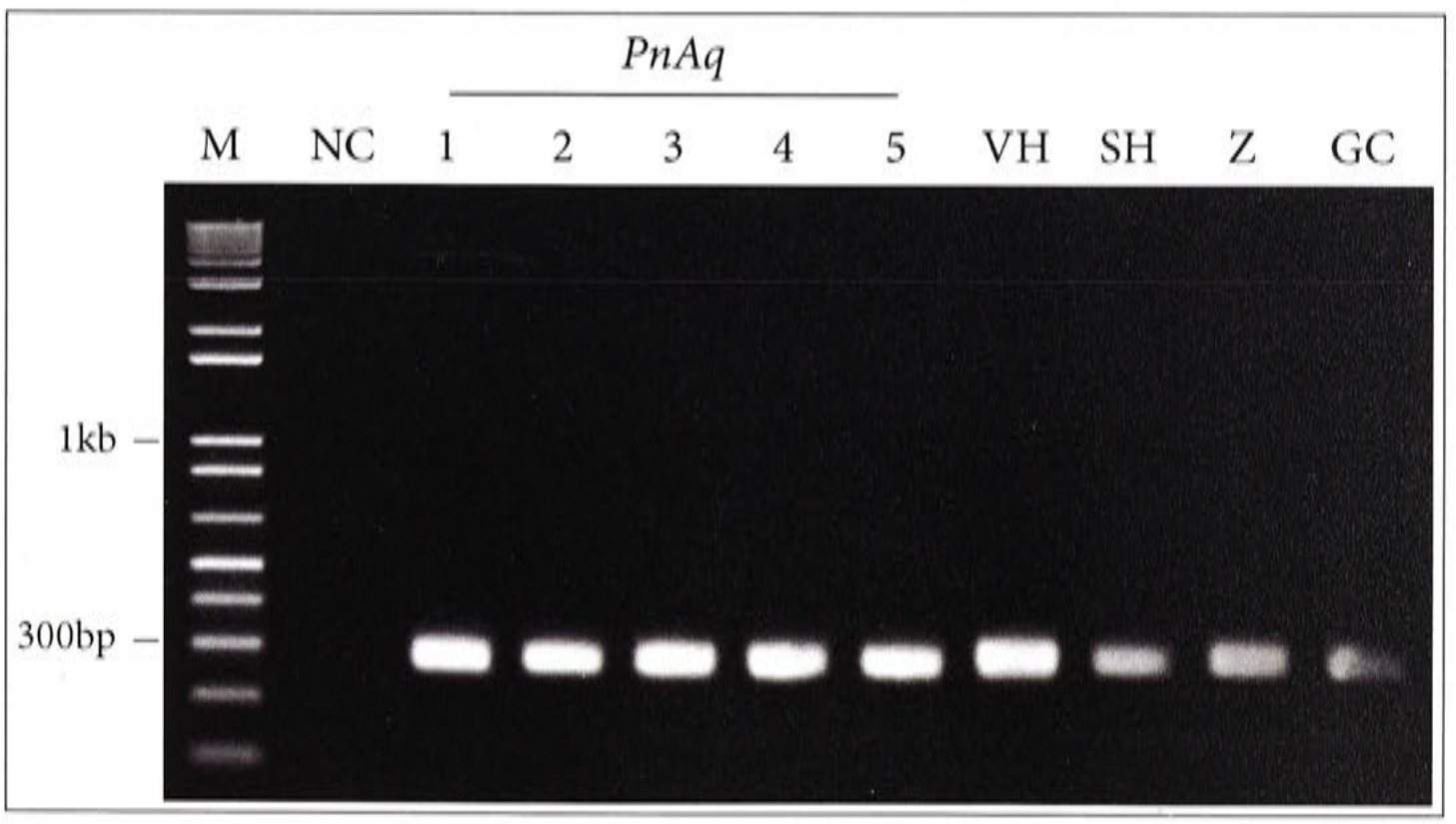


Figure 4.26 EtBr stained agarose gel of PCR products for SSCP analysis. Approximately equal amounts of DNA were loaded after concentration estimation by spectrophotometry. M–Marker, NC–negative control, VH–vegetative hyphae, SH–sporulating hyphae, Z–zoospores, GC–germinated cysts

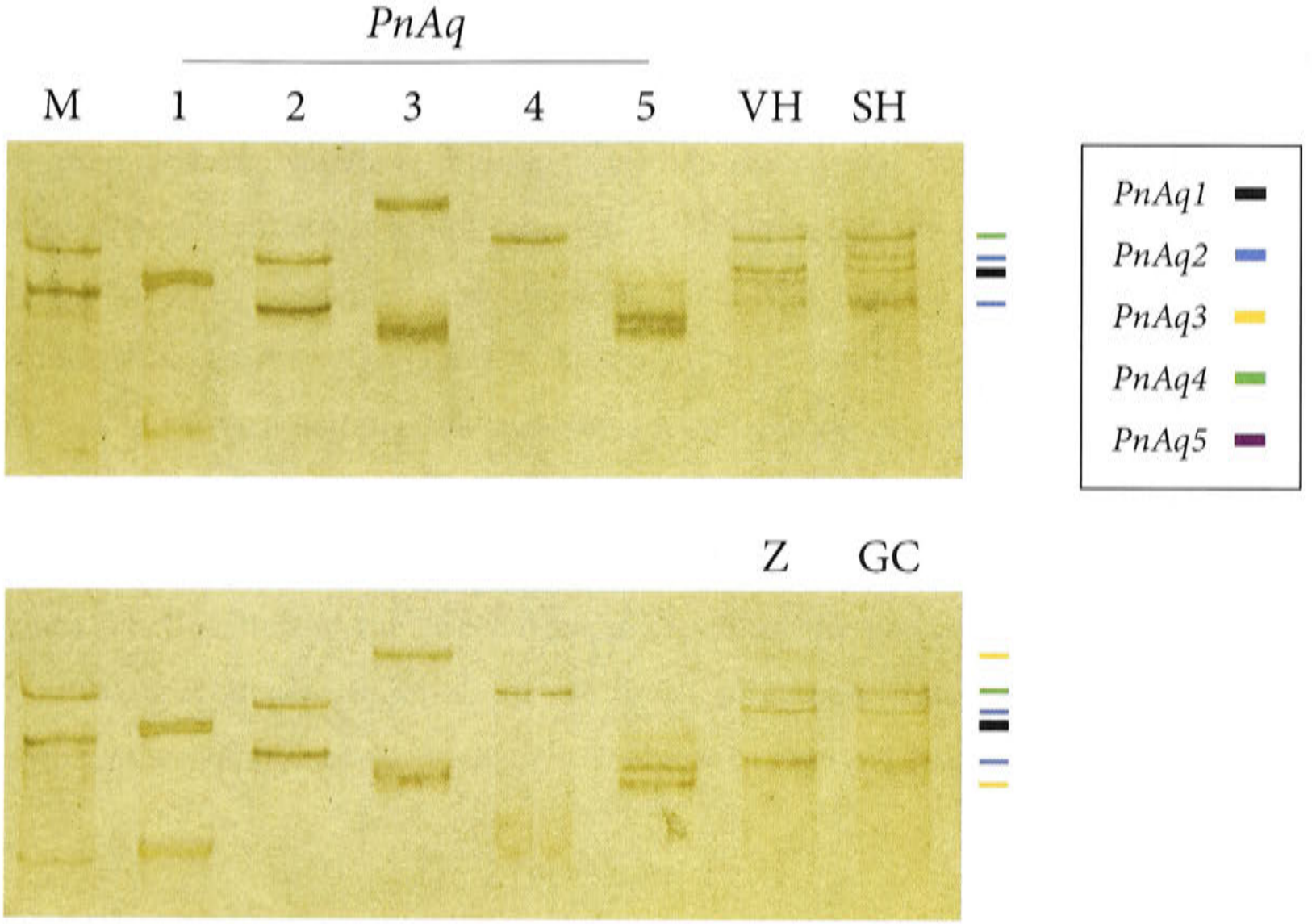


Figure 4.27 Silver stained SSCP gels of PCR products from *P. nicotianae* aquaporin reference genes and samples of all four developmental stages. For suggested gene expression pattern see Table 4.15 below. VH- vegetative hyphae, SH- sporulating hyphae, Z-zoospores, GC-germinated cysts

Table 4.15 Suggested expression pattern of *PnAq* genes as inferred from SSCP analysis

	PnAq				
	1	*2*	*3*	*4*	*5*
VH	x	x	-	x	-
SH	x	x	-	x	-
Z	-	x	x	x	-
GC	x	x	-	x	-

4.2.6 Peptide design for production of antibodies

In order to generate antibodies against PnAq proteins, peptides were designed for use in immunisations. Potential antigenicity sites (predicted by WebAngis), hydropathy plots and predicted transmembrane loops (see above) as well as secondary structures, potential post-translation modification sites and conjugation methods required were considered when designing the peptides. The two peptide sequences chosen for immunisation are shown in Figure 4.28. Pep-Aq2 is a "Y-shaped peptide", a combined N-terminal and C-terminal peptide containing a cysteine-residue in the middle to facilitate conjugation to the carrier protein. This form of antigen enhances the chance of an immune response due to the fact that the peptide effectively displays two antigenic regions. The location of the selected peptides within the protein sequence is shown in the multiple sequence alignment of the five *P. nicotianae* aquaporin proteins and the graphic representation of the aquaporin structure in Figure 4.29 and Figure 4.30.

Pep-Aq1

Residues: DER have acidic or basic side chains (underlined)
Residues: VLCMIP have hydrophobic side chains **(bold)**
Residue: NQY has uncharged-polar side chains (*in italics)*

*

towards N terminus ◀ C **M** *Q Y Q N* **L** *N* **V I** D **P** *N* R E ▶ towards C terminus

Peptide analyses: 40% of the residues are hydrophobic, 20% are charged, and 40% are non-charged.

Pep-Aq2

Residues: HRE have acidic or basic side chains (underlined)
Residues: VCPLM have hydrophobic side chains **(bold)**
Residue: QS has uncharged-polar side chains (*in italics)*

*

towards N terminus ◀ **V** *Q S* **V** H **L** R C **V** E **M** H H **P** *Q* ▶ towards C terminus

Peptide analyses: 46% residues are hydrophobic, 34% are charged, and 20% are non-charged.

Figure 4.28 Peptide generation for the production of antibodies derived from the PnAq amino acid sequence. The cysteine (grey box) was added to facilitate conjugation to carrier protein. The asterisk (*) indicates the antigenic site as predicted by WebAngis.

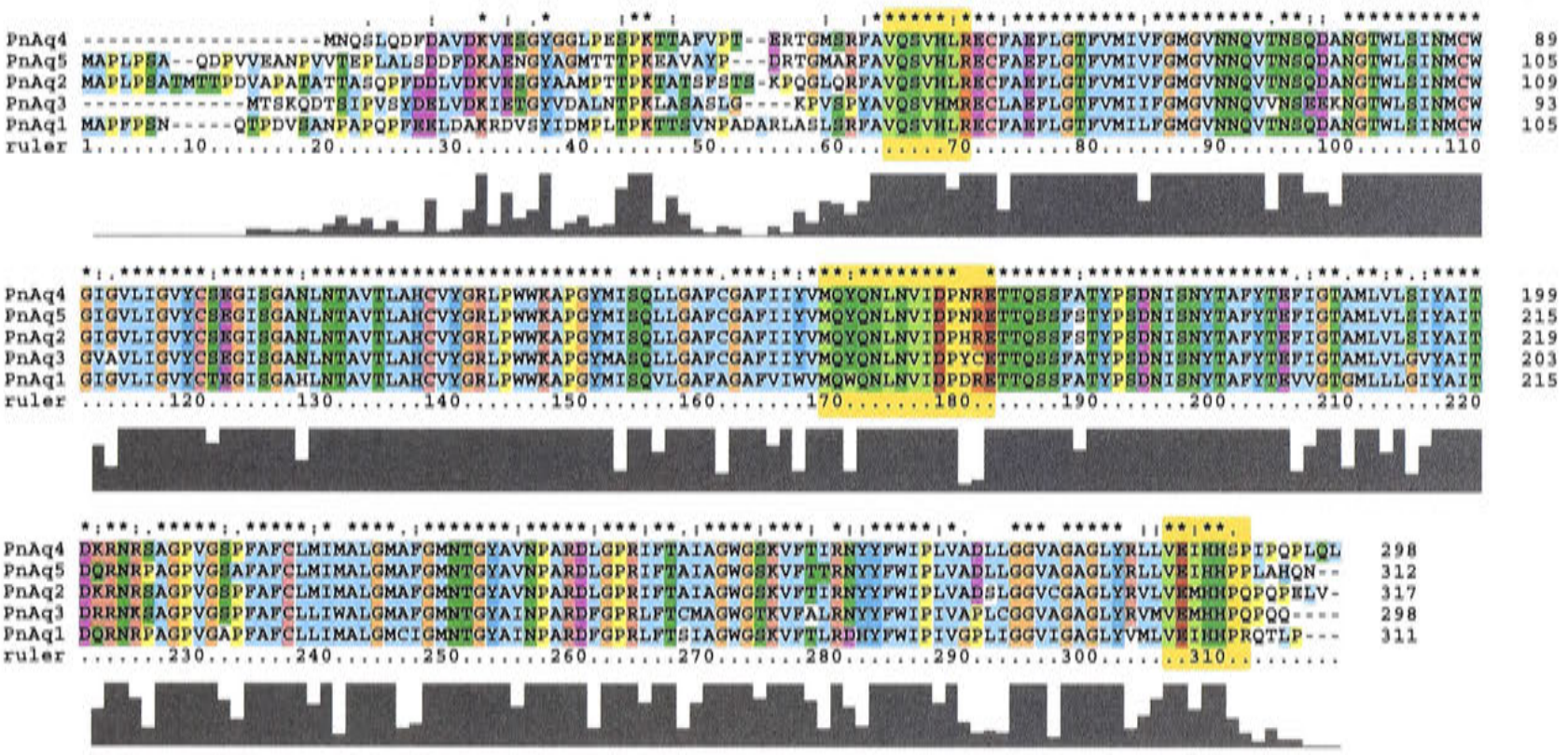


Figure 4.29 Multiple sequence alignment of *P. nicotianae* PnAq amino acid sequences. Highlighted regions show the selected peptides for antibody production.

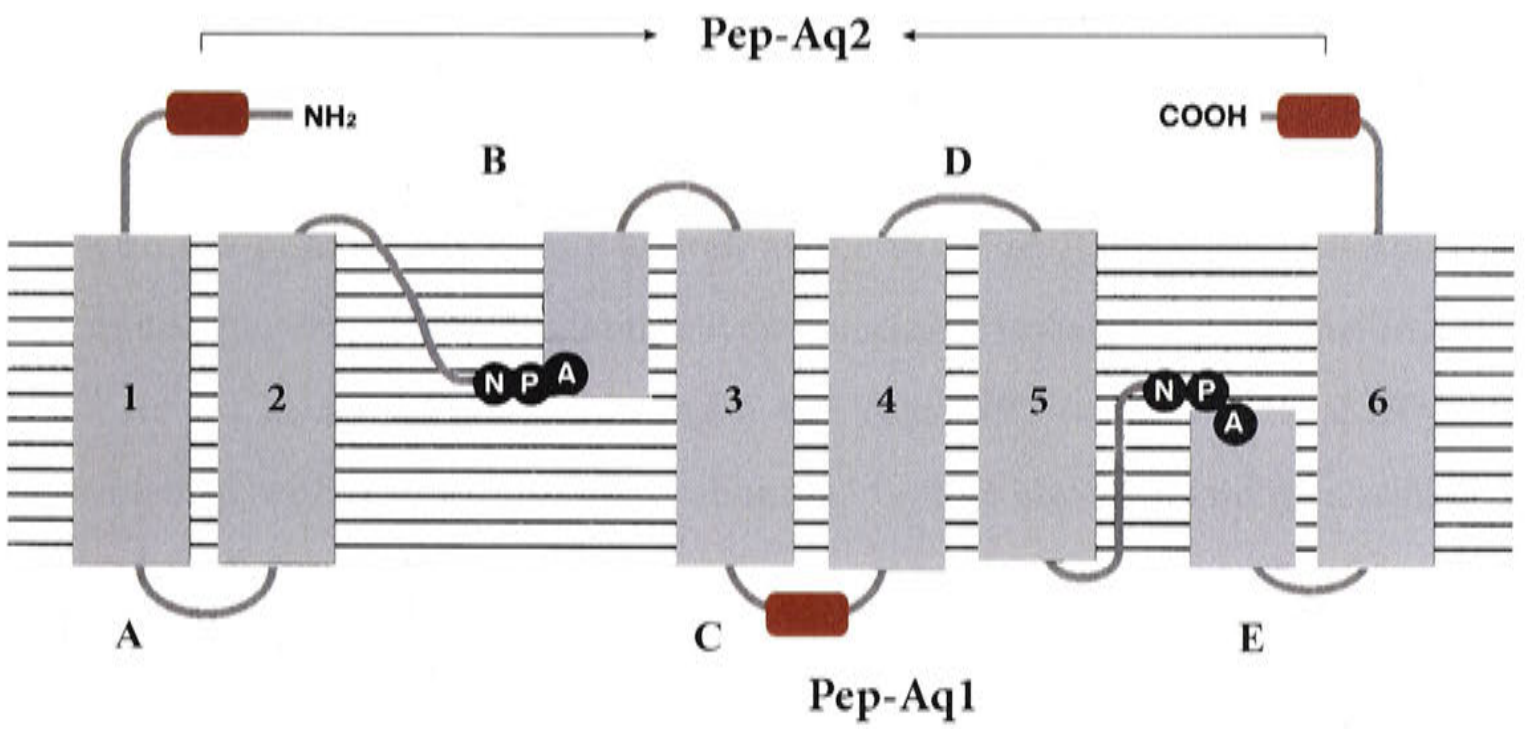


Figure 4.30 Localisation of peptides within the aquaporin protein. Highlighted regions represent peptide sequences Pep-Aq1 and Pep-Aq2 that were used for immunisation. (modified from Maeshima, 2001)

4.2.7 Antibodies against two PnAq peptides

For the production of polyclonal antibodies, two rabbits were immunised with both peptides Pep-Aq1 and Pep-Aq2 simultaneously (Rabbit #5 and #6) following protocols described in Section 2.5.4.1. Pre-immune sera were collected from both rabbits before the first immunisation. Rabbits were subjected to frequent booster immunisations and bleeds were taken at weeks 11, 13 and 20 and response to the peptides was monitored by ELISAs. Neither rabbit reacted towards peptide Pep-Aq1 after three booster injections, however, both rabbits showed a good response to peptide Pep-Aq2. Figure 4.31 shows the changes in antiserum titer in rabbit #5 over time. The antibody titer for the final sera was calculated by estimation of the inflection point of the post-immune graph and subsequent interpolation by drawing a line down the x-axis as depicted in Figure 4.32. The titer for the immune serum of both rabbits was estimated to be 1 in 1200.

Antibody purification was carried out following protocols described in Section 2.5.5 to increase specificity. Collected fractions were subjected to Bradford assays and SDS-PAGE and pooled according to protein levels and purity. The final protein concentration of both antisera was 0.4 mg/mL as determined by the Bradford assay. An SDS-PAGE loaded with 5 µg and 10 µg protein per lane and stained with Coomassie Brilliant Blue is shown in Figure 4.33.

In immunofluorescence assays, no labelling of the *Phytophthora* zoospores occurred with the pre-immune serum from either rabbit, while the immune sera from those rabbits reacted with the plasma membrane as well as the WEV of *P. nicotianae* zoospores fixed in a combined glutaraldehyde -formaldehyde solution. However, WEV labelling was not always observed in zoospores displaying a WEV in DIC images. Formaldehyde fixed zoospores tested with antibodies from rabbits #5 and #6 showed patchy labelling of the cytosol (data not shown).

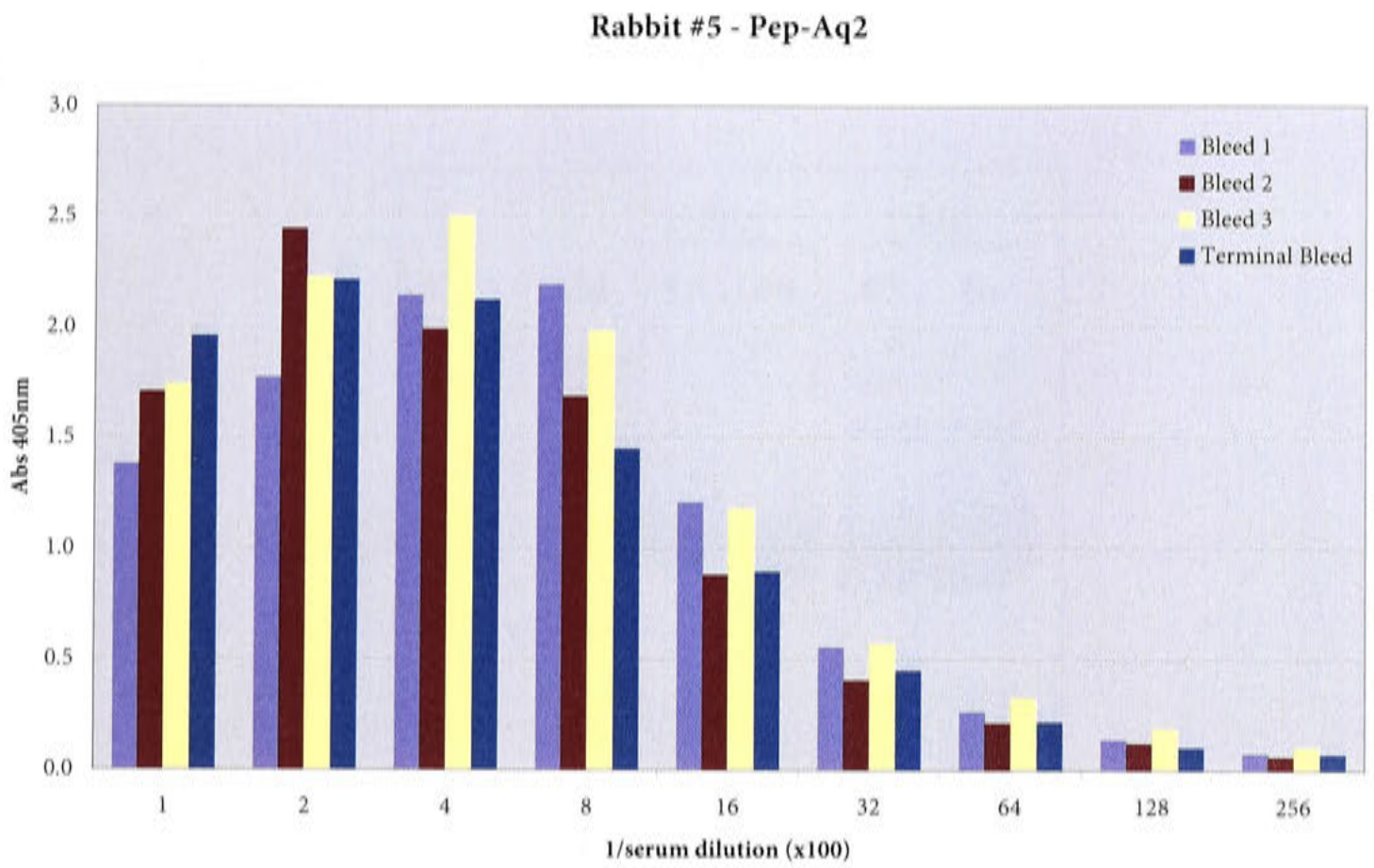


Figure 4.31 Antibody response to immunisation of rabbit #5 with peptide Pep-Aq2 as determined by ELISA. Serial dilutions of three bleeds during the immunisation process as well as the terminal bleed are shown. Depicted absorbance readings are after subtraction of pre-immune sera readings. Rabbit #5 reacted well to repeated booster injections with immunising peptide Pep-Aq2.

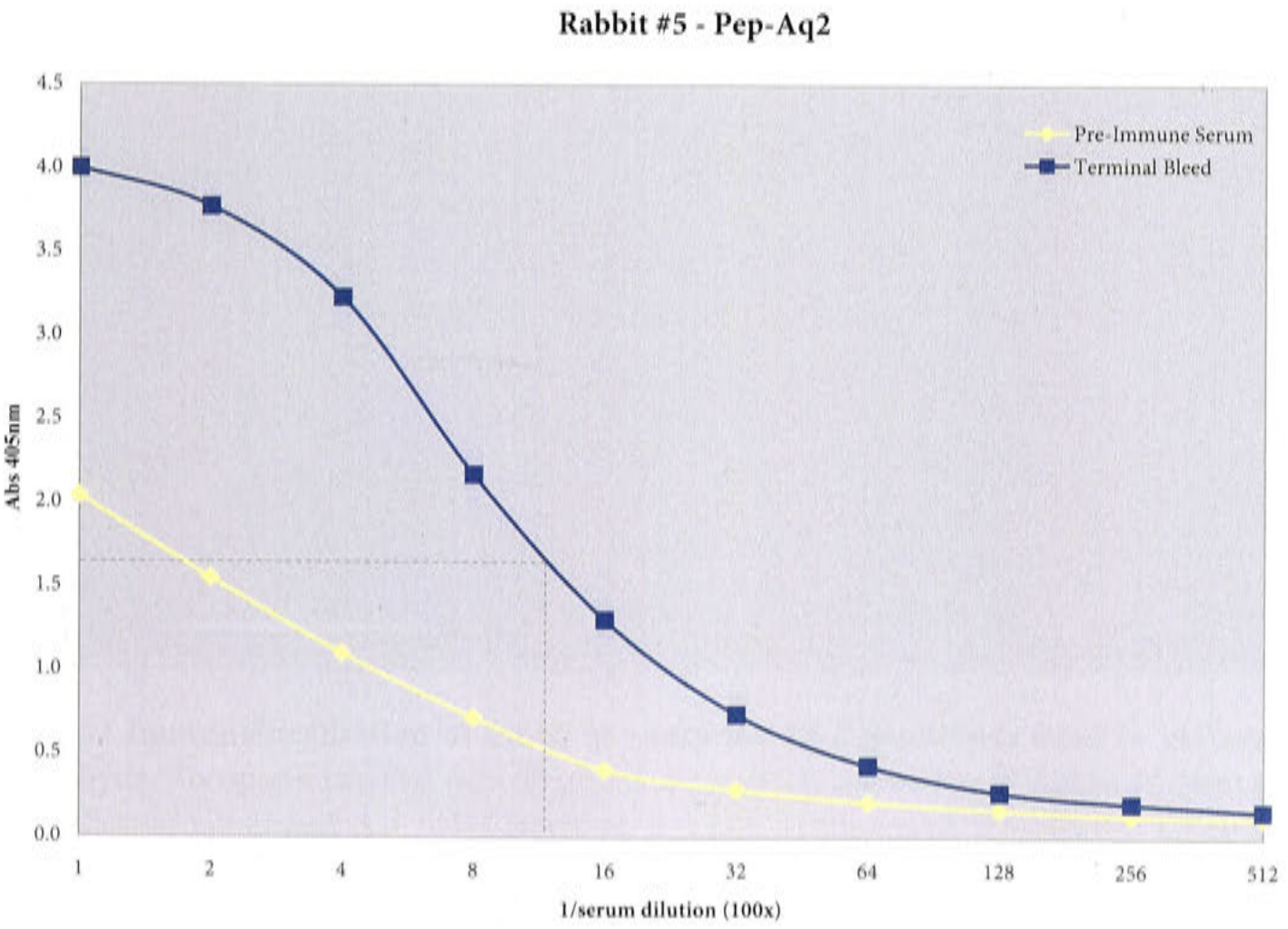


Figure 4.32 Antibody titer estimation of the rabbit #5 post-immune serum. The graph shows a plot of absorbance vs. antiserum dilution data obtained from ELISA analysis of the pre-immune serum (yellow) and the terminal bleed (blue). The grey, dotted line represents the estimated point of inflection of the post-immune graph. Drawing a line to the x-axis interpolates the titer to 1 in 1200.

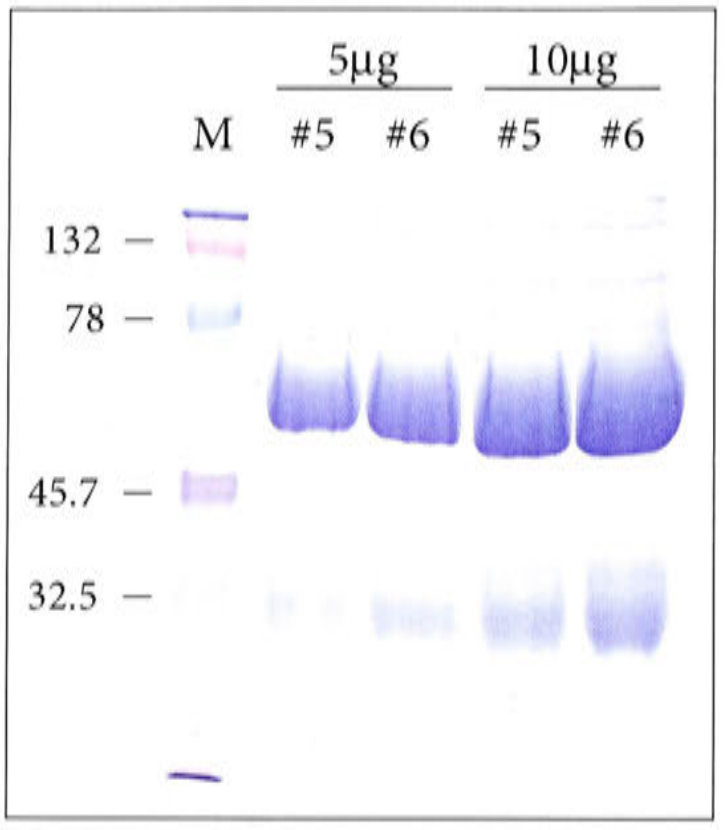


Figure 4.33 SDS-PAGE of purified antibodies after Coomassie Brilliant Blue staining. The final protein concentration after purification was estimated by Bradford analysis and 5 µg and 10 µg protein, respectively, were separated by SDS-PAGE. Numbers on the left indicate position of molecular weight markers in kDa.

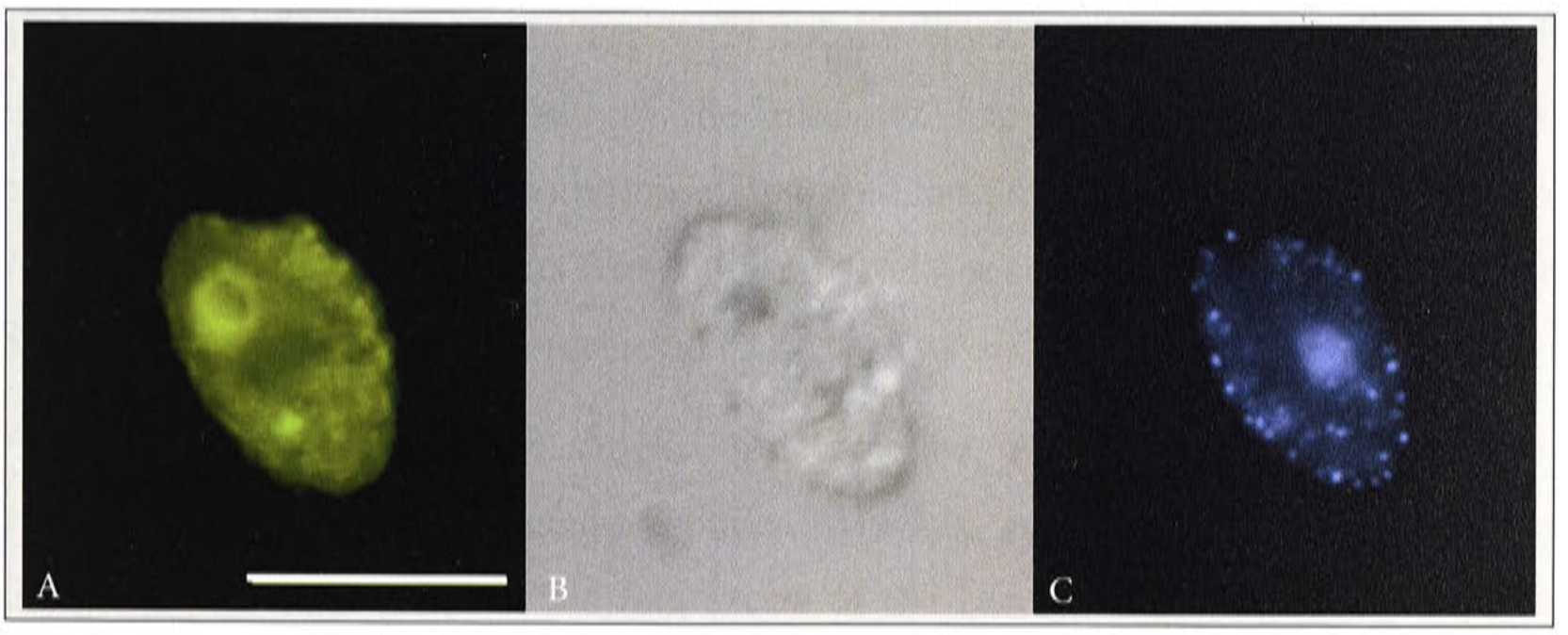


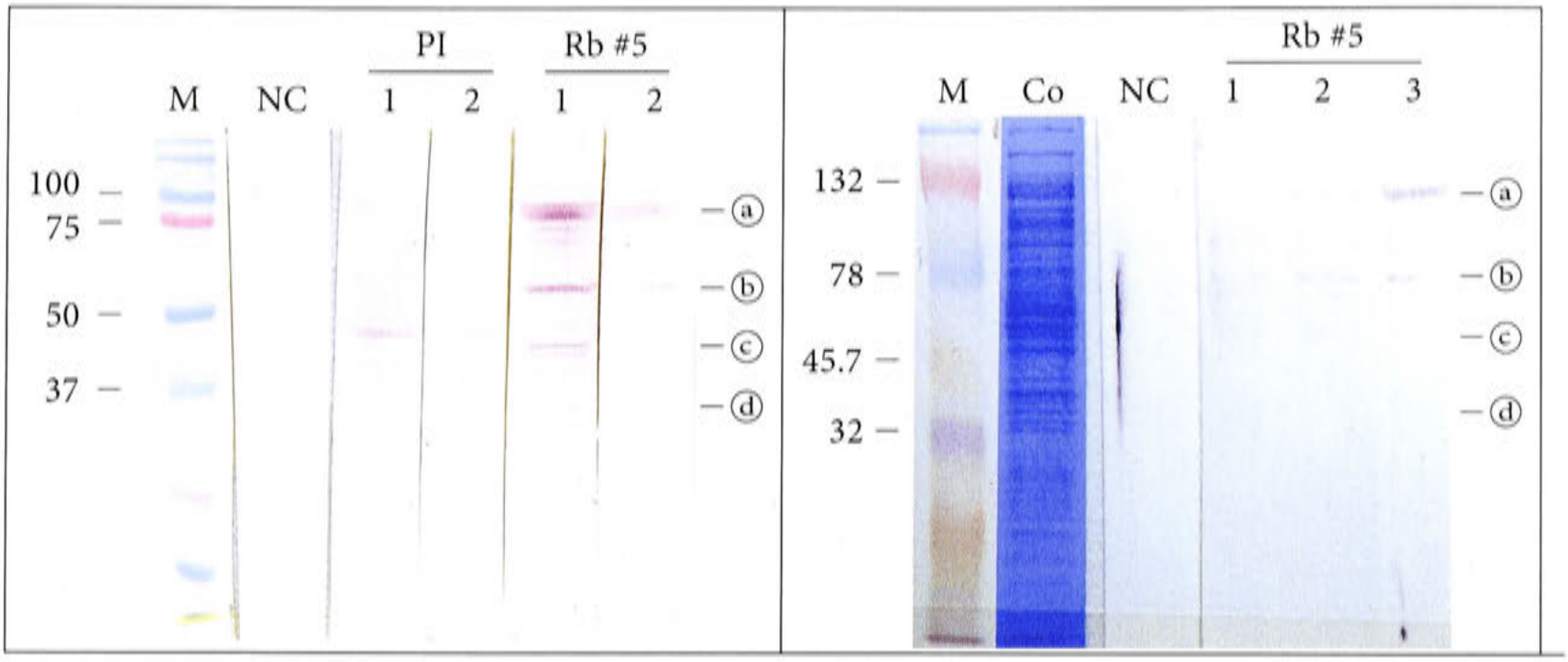
Figure 4.34 Immunolocalisation of PnAq in zoospores of *P. nicotianae* fixed in glutaraldehyde plus formaldehyde. Zoospores labelled with 30 µg/mL antipeptide-antibodies of Rabbit #5 react with the WEV and the plasma membrane. A - FITC labelling, B - DIC (Differential Interference Contrast), C - DAPI staining. Scale bar = 10µm

4.2.8 Western immunoblot analysis and peptide competition assays of polyclonal antibodies

Immunoblots were carried out on protein extracts from zoospores of *P. nicotianae* that had been solubilised in 2D-extraction buffer, separated by SDS-PAGE and blotted onto nitrocellulose membrane. Incubation with pre-immune sera of rabbits #5 and #6 resulted in faint labelling of two bands around 48 kDa and 100 kDa (rabbit #5) and many bands for rabbit #6. A dilution of 1 in 2000 of Rabbit #5 immune serum weakly labelled various bands in the range of 30 kDa to 100 kDa and four stronger bands at ~35/40 kDa, ~45/55 kDa, ~60/75 kDa and 110/120 kDa (Figure 4.35 a, Table 4.16). Estimated sizes for bands in blots varied depending on the protein marker used (Table 4.16). However, the strong 110/120 kDa band was also present in the pre-immune sera labelling and was therefore consider non-specific. Purification of antibodies slightly reduced the background labelling. Figure 4.35 b shows an immunoblot with purified antibodies of rabbit #5 using a concentration of 10 µg/mL to label different concentrations of zoospore proteins.

A peptide competition assay was carried out in order to determine whether the labelled bands were a result of specific binding of the antibody to the target protein or due to non-specific interactions. Pre-incubation of unpurified antibody with a 200-fold excess of the immunising peptide (i.e. Pep-Aq2) did not block labelling of any of the bands, suggesting non-specific binding of the antibody. Experiments were repeated with purified antibodies at a concentration of 10 µg/mL and increased peptide excess (up to 750-fold) using protein extracts from all four developmental stages of *P. nicotianae*. No decrease in signal intensity was detected (data not shown).

Peptide competition dot-blots showed, however, that immune sera of rabbit #5 and rabbit #6 recognise their respective peptides. Labelling with antibodies of rabbit #5 was successfully abolished with a 500-fold excess of the immunising peptide Pep-Aq2. No reduction in signal intensity was observed on *P. nicotianae* zoospore extracts (Figure 4.36). Antibodies of rabbit #6 reacted with Pep-Aq2, however, labelling was not abolished with a 1000-fold excess of peptide suggesting non-specific binding of the antibody (data not shown).



a) b)

Figure 4.35 Immunoblots of *P. nicotianae* zoospore protein extract after incubation with anti-peptide antibodies of rabbit #5. (a) Pre-immune and post-immune sera of rabbit #5 (lane 1 – 1:1000, lane 2 – 1:2000). Antibodies react with multiple bands in a range of 30 kDa to 100 kDa. The pre-immune serum labels two weak bands around 100 kDa and 48 kDa. (b) Purified antibody of rabbit #5. Antibody concentration of 10 µg/mL on 2.5 µg (lane 1), 5 µg (lane 2) and 10 µg zoospore protein (lane 3). Numbers on the left indicate position of molecular weight markers in kDa. Letters on the right indicate bands labelled. For further details see text and Table 4.16. M – Kaleidoscope marker, NC – 2nd antibody only, PI – pre-immune serum, Co – Coomassie staining

Table 4.16 Comparison of immunoblot results shown in Figure 4.35

Band	AB unpurified (Blot a - Kaleidoscope prestained standard)	AB purified (Blot b - Kaleidoscope Precision Plus)	PI
a	110 kDa	120 kDa	+
b	60 kDa	75 kDa	–
c	45 kDa	55 kDa	–
d	35 kDa	40 kDa	–

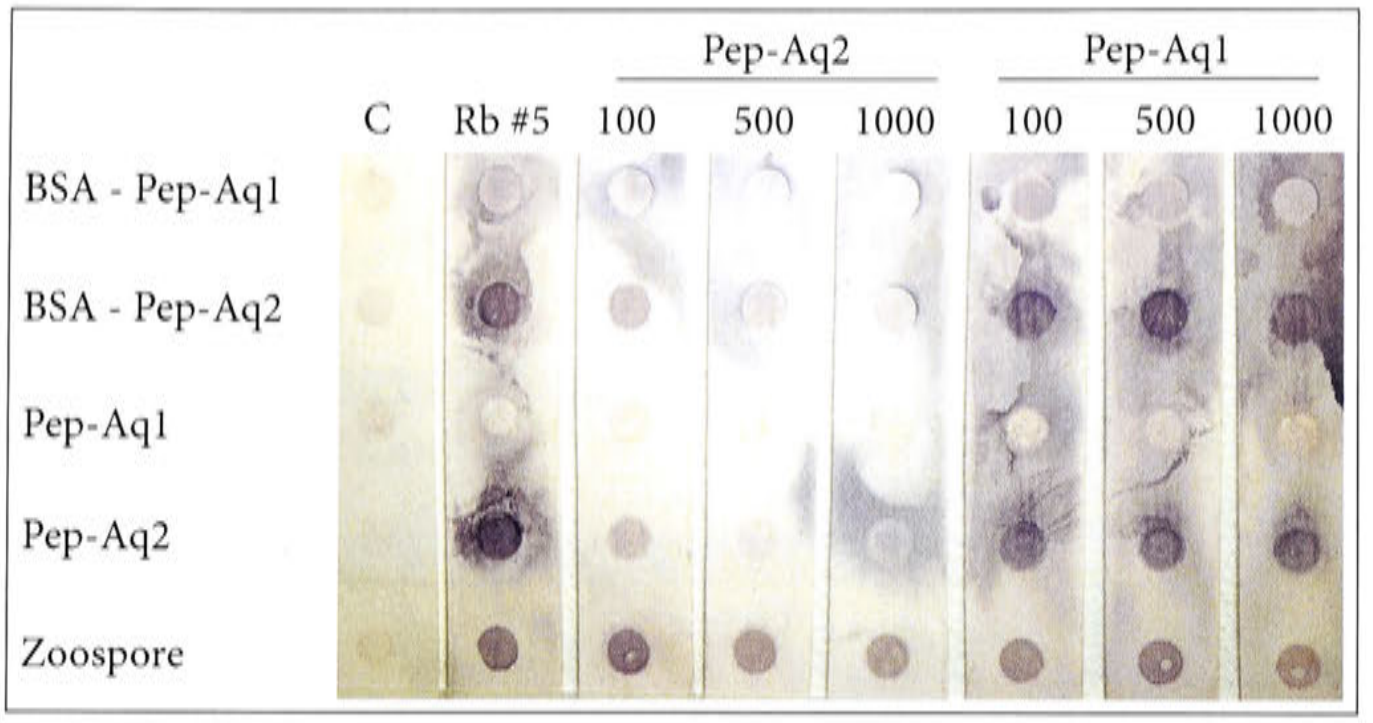


Figure 4.36 Immunodot blot of peptide competition assay of BSA-conjugated and unconjugated peptides labelled with immune serum from rabbit #5. Post-immune serum showed reduced antibody binding when incubated with the immunising peptide (Pep-Aq2), whereas the control peptide (Pep-Aq1) does not affect binding. Column 1 (C) displays 2nd antibody only labelling, Column 2 shows labelling with 10 µg/mL antibody from rabbit#5 only, whereas Columns 3 - 8 show reactions after pre-incubation of antibodies with a 100 - 1000-fold excess of peptides Pep-Aq2 and Pep-Aq1. Antibody binding is abolished by incubation with a 500-fold excess of peptide Pep-Aq2 (row Pep-Aq2 / column 4). No reduction in signal intensity was observed using peptide Pep-Aq1. Furthermore, the signal in zoospore proteins was not abolished by peptide competition.

4.2.9 Construction of a dsRNA construct for transformation experiments

A dsRNA construct expressing a 357 bp fragment of the *PnAq2* gene was cloned into vector pTEP5 and then transferred into TH210 for the use in *P. nicotianae* transformation experiments. The cloning procedure is described in the flow-chart shown in Figure 3.42. To obtain a plasmid that expresses dsRNA duplexes, a partial cDNA sequence of the *PnAq2* gene as well as a modified Kan^R gene was cloned into vector pTEP5. Both, KanR and the dsRNA sequence were PCR amplified using primers KANF/KANR and T3/Aq_dsRNA rev, respectively. A triple ligation was performed using the linearised vector pTEP5 (*Spe*I), the *Cla*I digested Kan^R PCR product and the double-digested (*Spe*I/*Cla*I) *PnAq2* dsRNA insert. A schematic representation of the dsRNA construct in vector pTEP5 is presented in Figure 4.37. Successful cloning and correct insertion of the inserts was confirmed by colony PCR, control digests of the isolated plasmid DNA, subsequent sequencing and alignment of the obtained sequence with the *PnAq2* gene (Figure 4.38).

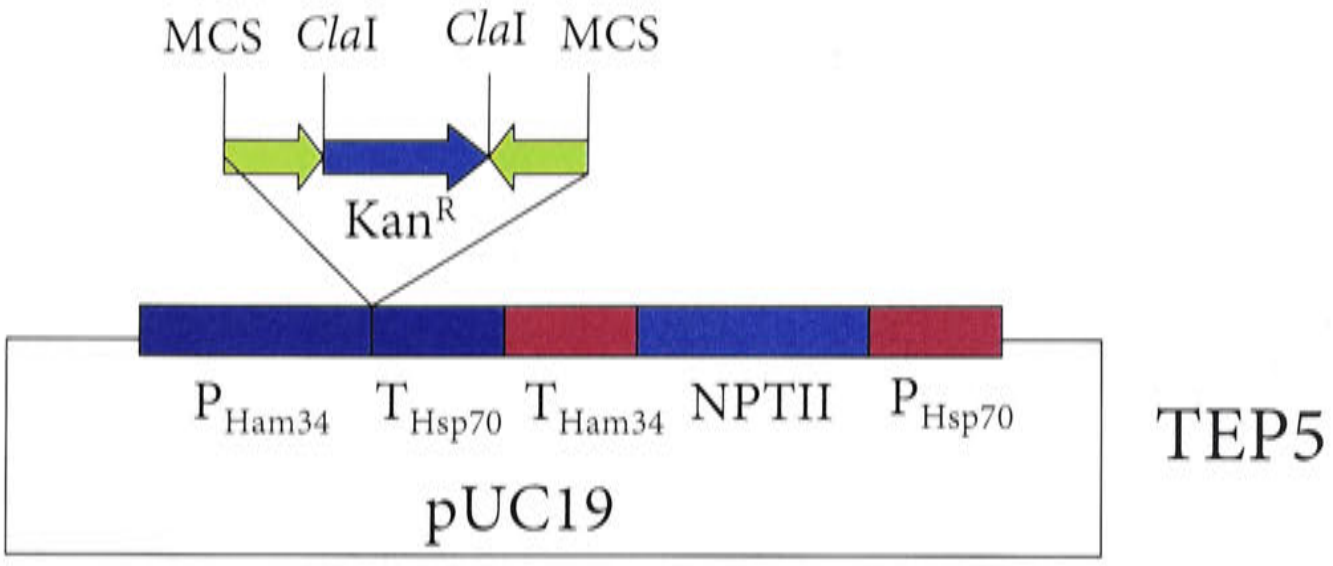


Figure 4.37 Map of vector pTEP5. Components of the vector are as follows: P_{HSP70} and P_{Ham34} – promoter fragments from the *hsp70* and *ham34* genes of *Bremia lactucae*, T_{Ham34} – terminator fragment from the *ham34* gene of *B. lactucae*, NPTII - neomycin phosphotransferase II gene. Insert: MCS – multiple cloning site. (Diagram by W. Shan)

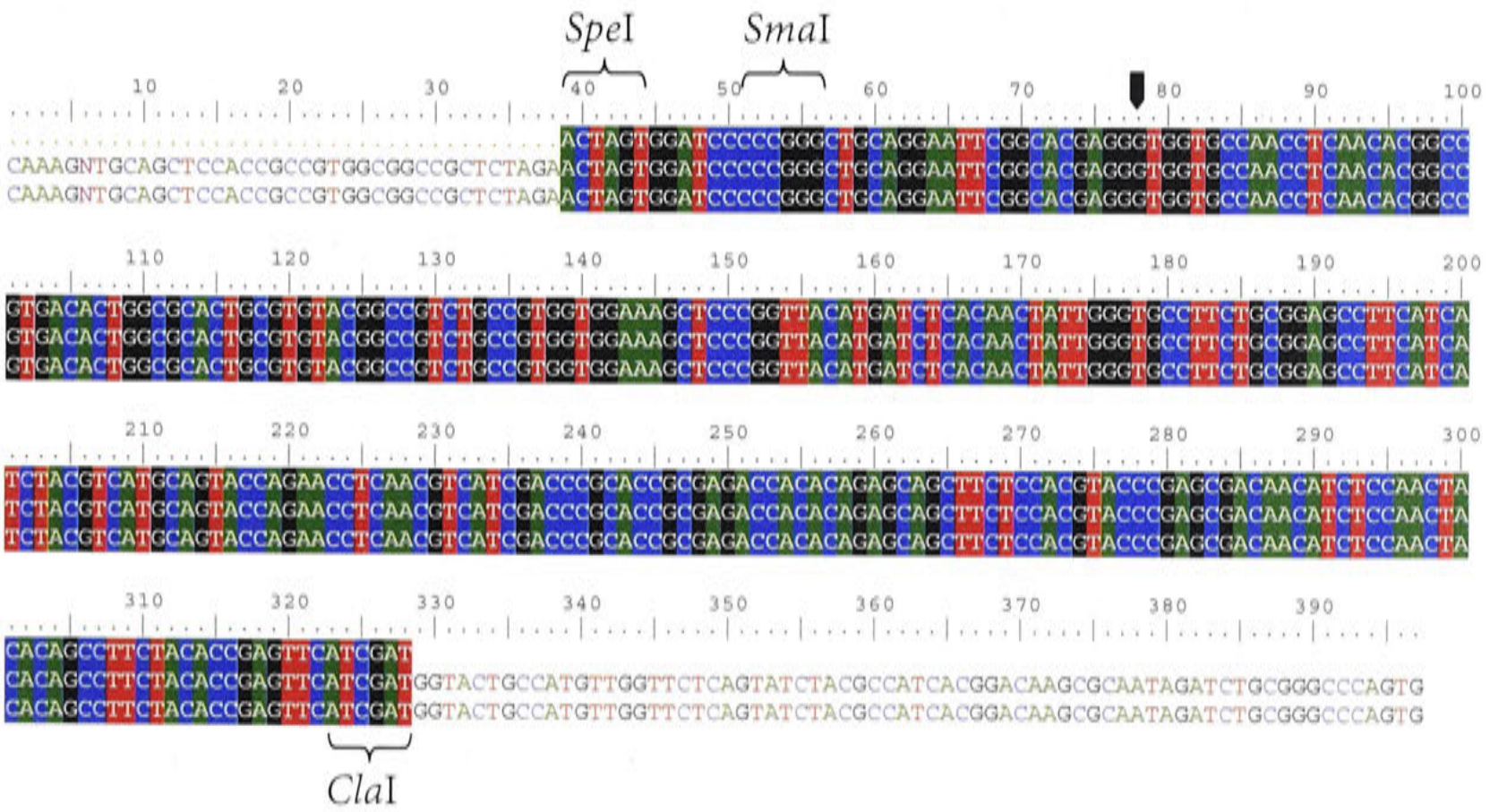


Figure 4.38 Alignment of the selected *PnAq2* dsRNA sequence with sequencing results of cloned plasmid. The dsRNA sequence (upper row) shows relevant 5' and 3' restriction enzyme sites obtained by using primer T3 (5') and a specifically designed GSP incorporating the *Cla*I site (3') for PCR amplification. The black arrow indicates the start of the actual *PnAq2* sequence. The alignment shows no mismatches between the predicted sequence and the sequencing data obtained from the plasmid.

4.3 Discussion

In this chapter, the cloning and characterisation of five aquaporin genes in *P. nicotianae* and their localisation in zoospores has been described.

Aquaporins constitute a large gene family with members found throughout all kingdoms of life. The idea of the existence of water transporting channels in cells was popular for many years (dating back to 1896 – see Benga, 2003), but it was only in the years 1988-1993 that the first aquaporin was successfully isolated and characterised (Denker *et al.*, 1988; Preston and Agre, 1991; Smith and Agre, 1991; Preston *et al.*, 1992). Since then, the number of publications on aquaporin-like proteins has grown exponentially. Research, especially functional characterisation, has focused predominantly on mammalian and plant aquaporins, with the exception of the *E. coli* aquaporins, and there are only a few studies available describing aquaporin-like proteins from other organisms such as fungi or protozoa. This study is the first report of the isolation and partial characterisation of aquaporin proteins in the oomycetes.

4.3.1 Cloning and structural organisation of the *PnAq* genes

Southern blot analyses suggested the presence of a small aquaporin gene family comprising at least four members in the *P. nicotianae* genome, based on *Bam*HI and *Sal*I digests of genomic DNA (Figure 4.13 a and Figure 4.16, respectively). Assembly of a genomic contig, covering the whole aquaporin locus, proved to be laborious because subcloning of several fragments from a number of restriction enzyme digests failed repeatedly. The analyses of numerous genomic subclones, however, showed that five and not four gene copies were present in the *P. nicotianae* genome, indicating that one band in both restriction enzyme digests, *Bam*HI and *Sal*I, represents two genes. To date, only 9 kb of approximately 12 kb of the sequence information obtained were assembled into a contig. It was not feasible to place the *Sal*I fragment S2, as no overlapping sequence information from other restriction enzyme digests was available. It is possible that the fragment is located at either side of the contig, however, based on expected sizes of the *Bam*HI and *Xho*I fragments that were not subcloned, it is highly likely that fragment S2 is situated upstream of the *PnAq2* gene. Unfortunately, attempts to close these gaps by sequencing aquaporin positive BAC clones were unsuccessful, most likely because of the repetitive nature of the aquaporin genes. Thus, the remaining *Bam*HI and *Xho*I fragments need to be subcloned and sequenced.

No introns were detected by comparison of the *PnAq2* cDNA clone (WS86) with the genomic clone within the region of overlap. In addition, full length ORFs with homology to aquaporin genes were detected in all *PnAq* genes, giving further evidence of the absence of introns in these genes. This is consistent with reports that the majority of genes in *Phytophthora* species do not contain any introns. In an analysis from 2002, it was shown that out of 63 *Phytophthora* genes only 21 genes (33%) contained, on average, 1.5 introns (Pieterse *et al.*, 1994). More recently, it was demonstrated that out of 33 *P. infestans* genes for which the full genome sequence was recovered, only eight genes (27%) had introns (six genes had one intron each and two genes had two introns each). Additionally, out of 68 partially covered genes, seven contained nine introns in total (Win *et al.*, 2006). Interestingly, mammalian and plant aquaporin genes generally contain two to five introns (e.g. Quigley *et al.*, 2001; Morishita *et al.*, 2004), whereas protozoan (Beitz, 2005) and yeast aquaporins do not usually have introns.

Of the five *P. nicotianae* genes, three displayed ambiguous in-frame translation start sites: *PnAq1* and *PnAq2* have two alternative ATG start sites and *PnAq5* displays three possible start codons. Promoters of all five genes were analysed for potential transcription factor binding sites in the 5' UTR. A TATA-box, a canonical motif commonly found within 35 bp of the TSS in most eukaryotic genes, which is, however, not always present in *Phytophthora* genes, was detected in two genes (*PnAq4* and *PnAq5*). Other promoter elements which regulate gene expression, like CAAT motifs and CT-rich regions, are in general located within 200 bp of the TSS (Lodish, 2000). A CAAT motif was found in the 5' flanking region of *PnAq2-4*, and all genes displayed CT-rich elements. A conserved 19 bp motif with the core sequence GCYCATTYYNCAWTTTNYY, found in many oomycete genes and usually situated within 100 bp upstream of the translational start site (Pieterse *et al.*, 1994; McLeod *et al.*, 2004), was found in all five promoter regions. Based on the detection of the promoter core elements and their locations in respect to the ambiguous translational start codons of three genes, it was concluded that the most likely start site of translation in *PnAq1* and *PnAq2* is the first in-frame ATG (Figure 4.18, position -102 and -117, respectively) and the central ATG at position -111 in *PnAq5*.

Sequences resembling AT-rich polyadenylation signals (consensus sequence AATAAA) are commonly found downstream of oomycete ORFs (Pieterse *et al.*, 1994; Kamoun,

2003; Win *et al.*, 2006). Putative polyadenylation signals, conforming with the signature sequence, were observed after the stop codon in the 3' UTR in all five genes.

4.3.2 Expression of the *PnAq* genes in *P. nicotianae*

RNA blot analysis using cDNA probe WS86, encoding for PnAq2, revealed a ~1.4 kb transcript in vegetative hyphae and germinated cysts, with the level of expression being highly upregulated in the latter cells. Expression was not detected in sporulating hyphae or zoospores (Shan *et al.*, 2004). Cross-reactivity of the probe with the other four genes in northern blot analysis is expected, as those genes show 77-93% homology to the cDNA probe used. Furthermore, successful hybridisation and cross-reactivity was demonstrated by Southern blot analysis. However, in genomic DNA blots genes are equally represented and the same amount of target DNA of each gene is presented to the probe, whereas in RNA blots the numbers of individual mRNA transcripts can vary significantly. Therefore, the detection of mRNA might be limited to *PnAq2* in northern blot analysis, with possible extension to the two other genes which display high homology to the probe's sequence (i.e. PnAq4 – 93% and PnAq5 – 89%). Sensitivity of RNA blots furthermore strongly depends on experimental conditions. Of particular interest are stringency conditions (i.e. hybridising and washing), which depend on time, temperature as well as salt and SDS concentration in the buffer. Conditions that are too stringent might not allow the detection of low abundance transcripts of aquaporin genes which exhibit low homology to the probe. Additionally, RNA blots were routinely subjected to several rounds of stripping in our laboratory in order to re-hybridise the blots with different probes. The harsh conditions used for stripping (e.g. washing of the blot in heated 0.1% SDS) not only remove the unwanted probe but are also likely to remove small amounts of RNA. Thus, sensitivity of RNA blots decreases each time the blot is stripped and low abundance target RNA might be lost during the procedure. Therefore, in order to determine whether aquaporin genes other than *PnAq2* are expressed in any of the four life cycle stages of *P. nicotianae*, a modified SSCP approach was performed.

All aquaporin genes were individually subcloned into the pBS vector by excision of fragments from two genomic subclones. *PnAq2* did not require subcloning, as it was the sole gene in one of the genomic subclones. Primers for the amplification of a PCR product were carefully chosen based on a multiple sequence alignment of all five

aquaporin DNA sequences. The fact that SSCP analysis is most sensitive (i.e. shows the best resolution) when small PCR fragments between 150 and 300 bp are used was also taken into consideration. Two highly homologous regions that only showed nucleotide substitutions in few positions were chosen as primer sites (SSCP FOR: two changes in *PnAq1*, one change in *PnAq1* and *PnAq3* in SSCP REV). This ensured amplification of a 274 bp PCR product from the individual genes. Repeated SSCP runs of PCR products obtained from individual genes were performed to optimise experimental conditions and achieve maximal resolution in order to distinguish genes by band pattern. A number of factors influence the separation of ssDNA on polyacrylamide gels, for example, gel concentration, running time, buffer system and temperature. It has also been reported that addition of some substances such as glycerol, when used in combination with TBE buffer (Kukita *et al.*, 1997) or the addition of PEG (Markoff *et al.*, 1997) to the gel matrix, enhances the resolution. PEG was not found to improve results in this study, however, the addition of glycerol did indeed enhance the resolution. The effects of glycerol on the resolution of SSCP analyses is due to a reduction in pH of the buffer, caused by the interaction of glycerol and borate ions (Kukita *et al.*, 1997). The result of the SSCP analysis from RNA of the four life cycle stages of *P. nicotianae* suggests that *PnAq2* and *PnAq4* are expressed in all four developmental stages, and that *PnAq1* is expressed in vegetative and sporulating hyphae, and germinated cysts. *PnAq3* is expressed in zoospores only. No expression of *PnAq5* was observed in any of the life cycle stages. As amounts of DNA samples loaded onto the SSCP gel were equalised after PCR and purification of the amplified product to ensure the visibility of bands, it is not possible to quantify the expression of individual genes. Therefore, the expression profile obtained through SSCP analysis does not contradict the northern blot result. RT-PCR is a highly sensitive technique and allows the exponential amplification of only a few mRNA transcripts. Therefore, this PCR-SSCP approach is considered more sensitive than northern blot analysis.

4.3.3 Characteristics of the PnAq proteins – conserved domains and their structure

The five *PnAq* genes encode proteins with homology to MIP-like proteins. A CDD search (Marchler-Bauer *et al.*, 2005) with the translated sequences from the genomic contig identified conserved domains with the MIP superfamily. This family includes, amongst others, human AQP2 and AQP3, bovine AQP0, *E. coli* AQPZ and GlpF, *A. thaliana* PIPs and TIPs, and *S. cerevisiae* Fps1. *P. nicotianae* protein sequences are 71-84% identical to each other and show 81-88% similarity. Analysis of the primary structure of PnAq2 with TMD prediction programs and a hydropathy plot showed that the protein is highly hydrophobic with a great number of residues predicted to be located within the membrane. All TMD programs used for analysis predicted highly similar overall topology with very comparable residue composition of the helix core regions. PnAqs feature the typical MIP structure with six TMD which are interconnected by loops A-E. The C- and N-terminus of the proteins are located in the cytoplasm. Aquaporins across all organisms analysed to date show striking sequence conservation in almost all TMDs as well as in loops B and E. The four motifs in TMDs 1-4 and the loop E signature are entirely conserved in all five PnAq proteins. Interestingly, even the variable positions within the motifs (as indicated by an x in the motif sequence in Table 4.10) are conserved within the *P. nicotianae* aquaporin group.

The glycine in TMD5 is thought to be important for the stabilisation of the pore structure as it interacts with another glycine in the conserved motif of TMD2 (Murata *et al.*, 2000; Fujiyoshi *et al.*, 2002). Zardoya (2005) reported, after analysis of approximately 450 MIP sequences, that the glycine in TMD 5 is generally missing in fungal GLPs and AQP11/12, and is highly divergent in insect proteins. In TMD 6, the glycine of the motif W...PxxG is replaced by an alanine in all *P. nicotianae* proteins except PnAq1. The signature motif in loop B (SGxHxNPAVT) shows probably the most significant changes. The histidine in the fourth position is only present in PnAq1, whereas all other *P. nicotianae* aquaporins show a substitution with asparagine. This residue is also thought to play a role in pore stabilisation (i.e. anchoring of the position of pore loop B) by forming an ion pair with the conserved glutamic acid in TMD 1 (Murata *et al.*, 2000). A change in this position has been observed in some fungal MIPs (e.g. *S. cerevisiae*, *Aspergillus nidulans,* and *Ustilago maydis* (Montalvetti *et al.*, 2004; Ishibashi, 2006),

however, there have been no reports on the relevance of protein function in regards to this particular residue to date. Changes in this position have furthermore been noted, for example in plant SIPs, some proteins from *Caenorhabditis elegans*, AQP11 and AQP12, however, it was shown that the whole signature motif is highly divergent in these proteins (see below, Ishibashi, 2006).

NPA-boxes are highly conserved signature motifs of the aquaporin protein family which are located in loops B and E. These loops, held together by interactions between the proline residues, fold back into the membrane where they form an integral part of the protein's pore (Murata *et al.*, 2000). In the PnAq proteins, the second NPA motif is completely conserved, whereas the second position of the NPA-box in loop B, proline, is replaced by threonine in all five proteins. Deviated NPA-boxes in loop B are extremely rare. It was shown by sequence analysis that only 20 out of 463 MIPs (4%) display non-conservative NPA-boxes in loop B (Zardoya, 2005). The number of proteins showing a change in the second position of the motif is narrowed down even further, as the majority of replacements occur in the third position of the motif (e.g. alanine to valine, threonine, serine or leucine). All mammalian AQP7 proteins were reported to have an alanine instead of the proline in the second position (Zardoya, 2005), however, the function of these proteins does not seem to be impaired. The transport of water, glycerol, urea (Ishibashi *et al.*, 1997; Kishida *et al.*, 2001) and even arsenite has been reported for this aquaglyceroporin (Liu *et al.*, 2002). Other organisms, shown to contain proteins with a non-canonical first NPA-box are *T. brucei*, TbAQP2 with asparagine-serine-alanine (NSA) (Uzcategui *et al.*, 2004), and a number of *Plasmodium* species which show asparagine-leucine/phenylalanine-alanine (N(L/F)A) (Hansen *et al.*, 2002; Beitz, 2005). TbAQP2, one of three aquaporin proteins in *T. brucei* and PfAQP from *P. falciparum*, showed permeation of water, glycerol and urea in *Xenopus* expression assays (Hansen *et al.*, 2002; Uzcategui *et al.*, 2004). These results strongly indicate that a change in the proline position of the NPA box does not necessarily lead to a major change in the protein's structure and its solute transport potential. It is possible that the substitution residue itself is capable of sufficient interaction with the other proline in the second NPA box, and thus does not change the structure. It is also conceivable, that other residues in the vicinity of the NPA-box compensate for the loss of the proline-proline interaction.

Recently, the introduction of a new aquaporin-like subfamily, the subcellular aquaporins, has been proposed. This group is characterised by highly divergent NPA-boxes in both loops, and it is suggested that proteins are not targeted to the plasma membrane but to membranes of intracellular components (Ishibashi, 2006). However, only a few members of this new group (mammalian AQP11, AQP12 and *A. thaliana* SIPs) have been functionally characterised and localised to date (Ishikawa *et al.*, 2005; Itoh *et al.*, 2005; Morishita *et al.*, 2005). As PnAq proteins only contain a slightly divergent NPA motif in loop B, they do not meet the criteria for subcellular aquaporins. Interestingly, no subcellular aquaporin-like protein sequences have been found in unicellular organisms (Ishibashi, 2006).

The narrowest constriction of the protein's pore, the selectivity filter which is thought to play a role in substrate selectivity, is formed by four residues as inferred by the comparison of the 3D-structure of AQP1 and GlpF. In general, selectivity of MIP channels is suggested to be mostly dependent on the hydrophobicity and steric organisation of the pore-lining residues. In human AQP1, the selectivity filter is shaped by Phe56 (TMD 2), His180 (TMD 5), Arg195 (loop E) and Cys189 (loop E) (Murata *et al.*, 2000), whereas in GlpF from *E. coli* these positions are represented by Trp48, Gly191, Arg206 and Phe200 (Fu *et al.*, 2000). Three out of four residues in the *P. nicotianae* aquaporin proteins correspond to the GlpF residues, suggesting that PnAq proteins belong to the group of glycerol facilitator proteins and may allow the transport of glycerol, and potentially also water and other small solutes like urea. However, for the validation of this idea, expression analysis studies in the *Xenopus* oocytes system or functional complementation in yeast need to be performed.

4.3.4 Immunolocalisation of PnAq

The aim of this part of the study was to generate polyclonal antibodies against the *P. nicotianae* PnAq proteins to reveal the localisation of aquaporins in *P. nicotianae* zoospores. Selection of peptides for the immunisation of rabbits took into consideration a number of factors that are important for the production of high quality anti-peptide antibodies. The availability of potentially antigenic regions was limited due to the high hydrophobicity of the proteins and the great number of TMDs, and was furthermore diminished by the fact that parts of loops B and E also form short helices that fold back into the membrane. Pep-Aq1, located in the extracellular loop C, does not contain any

conserved motifs but the region shows good homology between the five PnAq proteins. Due to the lack of further suitable regions, the intracellularly located protein termini were chosen for the second peptide Pep-Aq2. The residues of both termini were combined in a single peptide, separated by a cysteine placed in the middle, which acts as a conjugation point to the carrier protein. This resulting 'Y-shaped peptide' offers two antigenic sites to the immune system of the inoculated animal, thus enhancing the possibility of antibody generation (Mimotopes, 2001). Immune response in rabbits was variable. Both rabbits (#5 and #6) were inoculated with Pep-Aq1 and Pep-Aq2 simultaneously to increase the immune reaction and the possibility to obtain antibodies that recognise the native protein and not only the peptide (Mimotopes, 2001). In ELISA experiments, no response was observed against Pep-Aq1 after several booster injections, indicating low antigenicity of the peptide. However, the Y-shaped Pep-Aq2 triggered immune response in both rabbits.

Immunoblots prepared from *P. nicotianae* zoospore proteins labelled with PnAq antibodies from rabbit #5 reacted with four bands at ~ 35 (40), 45 (55), 60 (75) and 110 (120) kDa, depending on the molecular weight marker used. However, incubation with pre-immune serum also resulted in moderately strong labelling of two bands. One corresponded to the major antiserum band ~110 kDa which was therefore considered non-specific. It has been shown, that the aquaporin tetramer, under certain circumstances, does not fully solubilise in SDS, and therefore up to four bands can be seen in SDS-PAGE and in immunoblots, representing each possibly oligomerisation state of the single subunit (Maeshima, 1992; Borgnia *et al.*, 1999). As the inferred molecular weight of all five proteins ranges from 32.6 to 34.4 kDa, the bands around 35 kDa and 60 kDa could reflect monomers and dimers of the proteins, respectively. However, the occurrence of the third band at 45 kDa cannot be explained by oligomerisation of the aquaporin monomer. To elucidate whether any of the labelled bands were specific, peptide competition assays were carried out. Antisera from both rabbits were able to recognise the immunising peptide Pep-Aq2, however, only labelling of antibodies from rabbit #5 was reduced by pre-incubation with a 100-fold and completely abolished with a 500-fold excess of its peptide as shown in immuno dot blots. Labelling of zoospore protein in dot blots did not show a decrease in signal intensity. Peptide competition assay using antisera from rabbit #5 on immunoblots with protein from all four developmental

stages of *P. nicotianae* did not result in diminishing bands. Therefore it was concluded that antibodies display non-specific labelling in immunoblot assays.

Immunolocalisation studies showed labelling of the plasma membrane and some WEVs in *P. nicotianae* zoospores fixed in formaldehyde plus glutaraldehyde. Cells fixed in formaldehyde alone showed patchy labelling of the cytosol. This latter pattern is presumably due to poor preservation of the membrane, as fixation in formaldehyde alone (i.e. without the addition of glutaraldehyde), causes degradation of the plasma membrane. Labelling of only some WEVs in cells fixed in formaldehyde plus glutaraldehyde is consistent with the fact that the bladder needs to be open to the external environment for labelling to take place. If this only occurs during the water expulsion phase of the WEV cycle, it would explain why some WEVs visible in bright field were not labelled by the antibody (Mitchell and Hardham, 1999). However, as antibodies were considered non-specific in immunoblot assays, localisation patterns obtained in immunofluorescence experiments should be viewed with care as they might also be a result of non-specific binding of the antibodies.

4.3.5 Are aquaporins involved in *P. nicotianae* osmoregulation?

Proteins of the MIP family appear to have a broad range of physiological roles depending on the organism and tissue within which they occur. In mammals, the distribution and expression of the 13 individual proteins varies significantly, and so does their physiological role, ranging, for example, from water retention in the kidneys to production of tears or sweat (see section 4.1.3, Table 4.4). In plants, an even higher number of MIP-like proteins has been discovered, with up to 35 gene copies present in a single plant genome. These MIP-like proteins fall into four aquaporin subfamilies: TIPs, NIPs, PIPs and SIPs. MIP-like proteins are expressed in different subcellular compartments and tissues at different developmental stages in response to environmental stimuli, and are thought to play an important role in growth and general plant-water relationships (summarised in Maurel and Chrispeels, 2001; Tyerman *et al.*, 2002). Specific physiological roles are, for example, involvement in stress response, osmoregulation, root water transport, stomatal opening and rehydration of dry seeds (Kaldenhoff and Fischer, 2006b). In contrast, microbial genomes only contain a maximum of five aquaporin genes (Beitz, 2005; Pettersson *et al.*, 2005; Zardoya, 2005). Microbial cells, unlike cells from higher organisms, are directly exposed to the often

quickly changing environment, thus they need mechanisms that allows them to rapidly adapt to changing conditions. The involvement of aquaporins in osmoregulation in microbes, however, remains highly controversial to date (for details see Section 4.1.5).

Osmoregulation in *Phytophthora* zoospores is thought to be mediated by the WEV. In this study, it was suggested that antibodies raised against a peptide derived from PnAq protein sequences label the WEV and plasma membrane of *P. nicotianae* zoospores fixed in formaldehyde plus glutaraldehyde. The anti-aquaporin antibody thus displays the same labelling pattern as observed with the V-PPase antibody described in Chapter 3, Section 3.2.11. In *C. reinhardtii*, involvement of the glycerol facilitator CrMIP1 in osmoregulation and its localisation to the CV has been proposed (Anderca *et al.*, 2004). The only study to date with respect to CV localisation of an aquaporin was carried out by Montalvetti *et al.* (2004). The *T. cruzi* aquaporin TcAQP was shown to be a water channel only that co-localises with the V-PPase to acidocalcisomes and the CV. As both organelles are thought to be involved in osmoregulatory processes, it can be inferred that aquaporins play a role in osmotic adaptation in *T. cruzi*. Interestingly, comparison of the localisation of V-PPase and aquaporins in *P. falciparum*, carried out independently by two different research groups, also suggests the co-localisation of the two proteins (McIntosh *et al.*, 2001; Hansen *et al.*, 2002). However, as a CV has not been observed in *Plasmodium* species, it is likely that the proteins are located in the acidocalcisomes.

The amino acid residues that form the selectivity filter (as inferred from sequence comparison), suggest that the *P. nicotianae* aquaporins belong to the GLP cluster. In *S. cerevisiae*, it has been shown that the MIP-like protein Fps1, which transports glycerol and some other solutes but not water, is involved in osmoregulation (reviewed in Hohmann *et al.*, 2000; Kayingo *et al.*, 2001; Pettersson *et al.*, 2005). The accumulation of so-called compatible solutes is a common strategy in microbial osmoregulation (Yancey *et al.*, 1982). These solutes are accumulated under hyperosmotic conditions and released under hypoosmotic conditions (Poolman and Glaasker, 1998). In fungi, compatible solutes are usually polyols (e.g. glycerol or D-mannitol), whereas ions, sugars or amino acids are utilised by bacteria and protists (reviewed in Poolman and Glaasker, 1998; Hohmann *et al.*, 2000; Kayingo *et al.*, 2001). In yeast, the compatible solute is glycerol and its intracellular levels are controlled by the glycerol facilitator Fsp1 (Tamas *et al.*, 1999). This protein acts as a glycerol-exporter when cells are exposed to hypoosmotic

conditions, and Fsp1 deletion mutants have been shown to be highly sensitive to osmotic shock (reviewed in Kayingo *et al.*, 2001). Thus, it was concluded that Fsp1 is directly involved in the osmoregulation process in *S. cerevisiae*. In *P. nicotianae*, proline has been described as a compatible solute, with high levels detected in sporulating hyphae (Ambikapathy *et al.*, 2002). It was shown that hypoosmotic shock of *P. nicotianae* hyphae led to an approximately 50% decrease in free proline concentration within 30 min. (Ambikapathy *et al.*, 2002). Free proline levels within zoospores were extremely low, showing only 1% of the sporulating hyphae level. However, it was demonstrated that the expression of Δ^1-pyrroline-5-carboxylate reductase (P5CR), the enzyme that catalyses the final step in proline biosynthesis, is up-regulated in *P. nicotianae* zoospores, suggesting a high level of proline biosynthesis in zoospores (Ambikapathy *et al.*, 2002; Grenville-Briggs *et al.*, 2005). It was proposed that proline is expelled from the WEV during each water expulsion cycle, and has therefore to be restored continuously, which explains high levels of P5CR transcript but very low free proline levels in zoospores (Ambikapathy *et al.*, 2002). Another study examined levels of amino acid synthesis and levels of free amino acids in different life cycle stages of *P. infestans* (Grenville-Briggs *et al.*, 2005). Sixteen out of 21 amino acids analysed displayed a reduction of their freely available form of >80% when *in vitro* grown sporangia and zoospore values were compared, thus strengthening the idea that certain amino acids may act as compatible solutes in *Phytophthora* zoospores. Transport of amino acids by MIP-like proteins has not been reported, however, since several proteins show unusual permeation characteristics (e.g. anions for AQP6), it cannot be excluded that such transport mechanism exists. In accordance with the involvement of Fsp1 glycerol export in osmoregulation, PnAq proteins in *P. nicotianae* might control proline levels in zoospores. However, the existence of other osmoregulation relevant compatible solutes cannot be ruled out.

4.3.6 Gene silencing

As gene silencing experiments are a valuable approach to determine gene function *in vivo*, a dsRNA construct containing a 357 bp fragment of the *PnAq2* gene was cloned into the *Phytophthora* transformation vector pTH210, in order to carry out transformation experiments. However, due to the absence of protocols for stable

transformation of *P. nicotianae* at the time, it was not possible to perform these experiments.

A number of questions regarding the effectiveness and feasibility of experiments arise when considering the silencing of aquaporins in *P. nicotianae*. As it is likely that the dsRNA construct designed here would silence not only *PnAq2*, but also some of the other genes depending on homology levels between the dsRNA sequence and the target gene, it will be difficult to determine potential phenotypes and the assignment of locations and specific functions of each individual gene. The construction of gene-specific dsRNA constructs that do not interact with any of the other genes would be very difficult due to high overall sequence similarities between the five genes. Even if silencing of a single gene alone was accomplished, an observable phenotype might not be detected, as, depending on the functionality of each individual gene, other gene products might serve as back-up for the silenced gene. Even if multiple genes were silenced simultaneously, a phenotype might still not be observed. On the other hand, a phenotype observed due to multiple copy silencing would not allow for assignment of a specific function to a particular gene. Furthermore, it was shown that expression of one or more aquaporin genes of the group analysed in this chapter is up-regulated in germinated cysts, and that at least one of the corresponding proteins is located in the WEV and the plasma membrane of *P. nicotianae* zoospores. Aquaporins may be involved in potentially vital cell processes and silencing of the gene(s) might be fatal in any of the developmental stages. As transformants would have to pass through all life cycle stages to obtain germinated cysts for analyses, this could be a limiting factor.

In conclusion, this study presents the gene structure and gene expression of five PnAq genes from *P. nicotianae*. Proteins are predicted to belong to the MIP superfamily and are thought to transport glycerol and/or water. The exact role of individual genes in *P. nicotianae* has yet to be determined, however, a function in osmoregulation of at least some genes has been proposed. Further studies using the *Xenopus* oocyte expression system to determine functionality of genes, transformation of *P. nicotianae* with constructs leading to the expression of GFP-tagged proteins to determine the localisation of individual proteins and possibly gene silencing, may clarify the function of aquaporins in *P. nicotianae*.

Chapter 5

Phylogenetic analysis of V-PPase and aquaporins in *Phytophthora* species

Chapter 5: Phylogenetic analysis of V-PPase and aquaporins in *Phytophthora* species

5.1 Introduction

5.1.1 Databases

Databases (DBs) are large sets of information stored in a systematic way on a computer system and enable the efficient retrieval of data subsets upon request. With the rapid developments in computer technology over the past two decades, the storage and retrieval of specific data has become widely available to researchers. DBs catering for the specific needs of biological sciences have been created and in conjunction with bioinformatic programs are invaluable tools for researchers today. Advanced user interfaces allow for easy access and retrieval of the stored data by querying the DB, without a need to understand the underlying structure and complexity of the DB itself. A number of DBs rely heavily on the input of data by the user (e.g. Wikipedia, NCBI GenBank), whereas others are generally inaccessible to user modifications (such as PubMed, Web of Science).

5.1.1.1 The Phytophthor*a Genome Consortium* (PGC)

The PGC, formerly '*Phytophthora* Genome Initiative', was founded as a collaborative approach to unravel the mysteries of the destructive plant genus *Phytophthora*. Of particular interest here is the molecular basis of general host-pathogen interactions and the processes of infection and resistance by the pathogen and host, respectively. In 2004, the sequencing of the first two *Phytophthora* genomes (*P. sojae* and *P. ramorum*) was completed at the JGI. Additional sequence resources include a few BAC libraries and a large number of small and large-scale EST projects from a range of *Phytophthora* species (see Chapter 1). Sequence information obtained is processed with the help of the National Center for Genome Resources, using the analysis pipeline XGI (genome initiative for species X - http://www.ncgr.org/xgi). XGI allows researchers to collaboratively extend a single DB by submitting project-relevant sequences to the central server over the internet, where raw sequences then undergo an automated screening and DB implementation process (Figure 5.1). The screening process removes low quality reads as well as vector contamination, poly A/T stretches and common adapter or linker sequences. Further analyses (e.g. BLAST similarity searches, motif and

structure searches) are then carried out on the 'mature' and deposited sequence (Waugh *et al.*, 2000; Inman *et al.*, 2001; http://www.ncgr.org/xgi/).

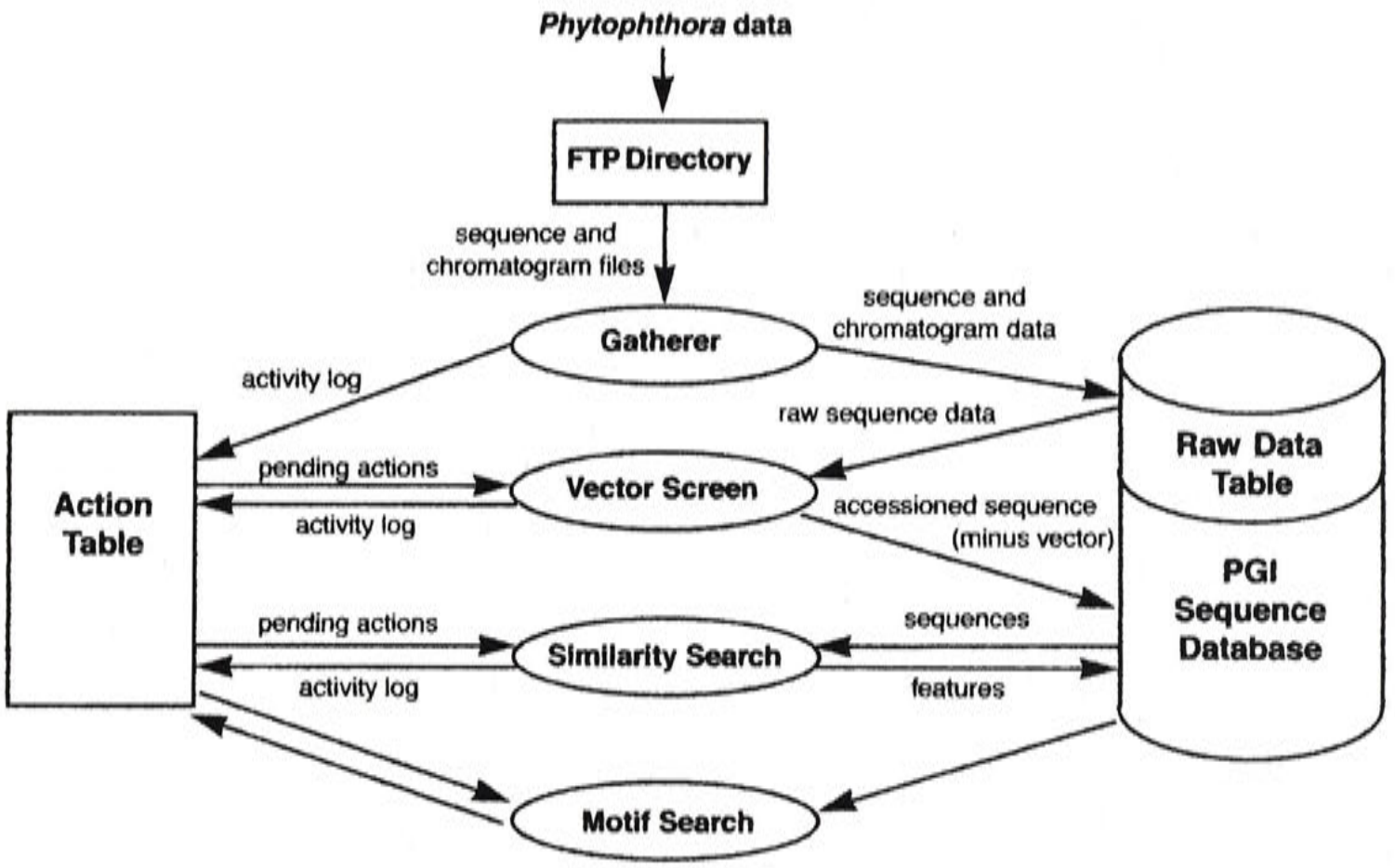


Figure 5.1 The PGC DB: Informatics and analysis for distributed pathogenomic research. (Source: Waugh *et al.*, 2000)

The DB of the PGC, a publicly accessible resource, is the PFGD, which contains transcript, genomic, gene expression and functional assay data of *P. infestans* and *P. sojae* (Gajendran *et al.*, 2006). It is composed of all available data from the PGC and the Syngenta *Phytophthora* Consortium (https://xgi.ncgr.org/spc/). Furthermore, all other publicly accessible *P. infestans* data coming through the XGI pipeline are included. PFGD is also linked to the 'Solanaceae Genomics Database', which contains expression study data of solanaceae response to *P. infestans*, therewith giving access to host-pathogen interaction data (Gajendran *et al.*, 2006).

Another resource, the 'VBI Microbial Database' (VMD) which was launched in April 2005, is designed to host microbial genome sequences, which are taken from different sources (Tripathy *et al.*, 2006). At present, it contains sequences and annotations of the *P. sojae* and *P. ramorum* genomes (genomic and EST data), as well as the genomes of *Alternaria brassicicola* and *A. tumefaciens*. VMD not only contains highly sophisticated browsers and query pages which give easy access to the data, but also a number of toolkits that are available to the end-user (e.g. programs for multiple sequence

alignments and the validation of gene models on the fly, EMBOSS suite with ~160 sequence analysis tools), which allow for easy evaluation of data on the spot (Tripathy *et al.*, 2006). In addition, VMD features an annotation tool, which enables researchers to actively participate in the annotation process of the genomes through a web interface. Gene models are predicted by four different algorithms at JGI, and only the best model is stored in VMD after validation (see Tripathy *et al.*, 2006). However, as the genome assembly is still incorrect in places and predicted gene models are not necessarily accurate, manual verification and annotation of models need to be performed. In 2004, an annotation jamboree was held, which led to the manual annotation of a number of genes.

All available *Phytophthora* DBs as well as a number of other DBs that may be useful in gene analysis are listed in Table 5.1.

In this study, the genomes of *P. sojae* and *P. ramorum* were searched for V-PPase and aquaporin genes (JGI). Furthermore, PFGD was mined for *P. infestans* ESTs.

Table 5.1 Useful databases (as of August 2006)

DB	Name	Description	URL
General DB			
GenBank		General DNA sequence DB	http://www.ncbi.nlm.nih.gov/
TIGR	The Institute for Genomic Research	Complete genome sequence DB	http://www.tigr.org/
Oomycete DB			
PFGD	Phytophthora Functional Genomics Database.	Genome/EST DB of *P. sojae* and *P. infestans*	http://www.pfgd.org/pfgd/
ESTAP	EST Analysis Pipeline	EST analysis platform	http://staff.vbi.vt.edu/estap/
JGI (*P. sojae)*	Joint Genome Institute	Genome sequencing project of *P. sojae*	http://genome.jgi-psf.org/sojae1/sojae1.home.html
JGI (*P. ramorum)*	Joint Genome Institute	Genome sequencing project of *P. ramorum*	http://genome.jgi-psf.org/ramorum1/ramorum1.home.html
VMD	VMD Microbial Database	*P. sojae* EST database	http://phytophthora.vbi.vt.edu/
OGD	Oomycete Genomics Database	EST DB (other than *Phytophthora*)	http://www.oomycete.org/ogd/
Other DB			
JGI (*T. pseudonana)*	Joint Genome Institute	Genome sequencing project of *T. pseudonana*	http://genome.jgi-psf.org/ramorum1/ramorum1.home.html
PlasmoDB	Plasmodium Genome Database	Genomic sequencing projects of *Plasmodium* spp.	http://www.plasmodb.org/plasmo/home.jsp
ApiDots	Apicomplaxan Database	EST sequences from numerous Apicomplexan parasites.	http://www.cbil.upenn.edu/apidots/
CryptoDB	Cryptosporidium Genome Database	Genomic and EST sequences of *C. parvum* and *C. hominis*	http://cryptodb.org/cryptodb/

5.1.2 Phylogeny

The word phylogenetics is derived from the Greek (phylon = tribe, race and genetikos = relative to birth, from genesis = birth) and describes a scientific discipline which deals with the reconstruction of evolutionary relationships among and between organisms (http://en.wikipedia.org/wiki/Phylogenetics). With the ever increasing amount of available genomic data, and the improvement of data retrieval by means of DB, in detail analysis and direct comparison of genomes from different species (i.e. comparative genomics) has become an important method in scientific research. Comparative genomics implements the use of a range of bioinformatic tools which enable the researcher to compare DNA or protein sequences and reconstruct the evolutionary history between these sequences. Comparing phylogenetic distances between organisms may help to not only understand general evolutionary history (divergence time) of organisms, but also aid in the prediction of function of novel genes by finding structurally important segments in proteins, detect orthology and paralogy between genomes and more specifically, help in mapping of pathogenic strain diversity (Olken, 2002; Holder and Lewis, 2003).

5.1.2.1 Phylogeny and the reconstruction of phylogenetic trees

Phylogenetic trees are now commonly reconstructed from either protein or DNA sequences. The actual analysis consists of a serious of steps: retrieval of data, multiple sequence alignment, phylogenetic estimation, tree building, estimation of the reliability of the tree and visualisation. A number of different algorithms/methods are available for reconstructing phylogenies, each displaying advantages and disadvantages (Figure 5.2). Details are beyond the scope of this thesis, however, a brief summary will be given below. For more details about different phylogenetic inference methods please refer to the excellent review papers by Holder and Lewis (2003) and Thornton and DeSalle (2000).

Traditional approaches for phylogenetic estimation are Maximum parsimony, Maximum Likelihood and the two distance-based methods UPGMA (Unweighted Pair Group Method using Arithmetic Averages) and NJ (Neighbour-Joining algorithm). In early 2000, a new approach (Bayesian method) with underlying Markov chain Monte Carlo methodology has emerged. The most frequently used method by researchers is the distance-based NJ method as its performance is relatively fast. Tree reliability is enhanced with a statistical method called 'bootstrapping', which uses pseudo-replicate

sets (usually between 500 and 1000) of the original data by random resampling of the characters. Tree reconstruction is carried out on all datasets, trees are compared and a consensus tree representing the most reliable tree reconstruction is computed. Bootstrap values (i.e. the proportion of trees/replicates in which a branching point is recovered – Soltis and Soltis, 2003) are given for each branching point of the tree and are used to evaluate the support of a particular grouping of organisms. However, interpretation of bootstrap values are blurry and controversial, as no clear cut-off values are defined for the support of strong/weak grouping, thus trees have always be viewed with caution (Holder and Lewis, 2003; Soltis and Soltis, 2003).

The methodology used in this chapter for the phylogenetic analysis of the V-PPase single gene family and the aquaporin multiple gene family in *Phytophthora* is described in Section 2.7 and also summarised in the 'DB manual' in Appendix H.

Method	Advantages	Disadvantages
Neighbour joining	Fast	Information is lost in compressing sequences into distances; reliable estimates of pairwise distances can be hard to obtain for divergent sequences
Parsimony	Fast enough for the analysis of hundreds of sequences; robust if branches are short (closely related sequences or dense sampling)	Can perform poorly if there is substantial variation in branch lengths
Minimum evolution	Uses models to correct for unseen changes	Distance corrections can break down when distances are large
Maximum likelihood	The likelihood fully captures what the data tell us about the phylogeny under a given model	Can be prohibitively slow (depending on the thoroughness of the search and access to computational resources)
Bayesian	Has a strong connection to the maximum likelihood method; might be a faster way to assess support for treesthan maximum likelihood bootstrapping	The prior distributions for parameters must be specified; it can be difficult to determine whether the Markov chain Monte Carlo (MCMC) approximation has run for long enough

Figure 5.2 Comparison of methods for phylogenetic tree construction. (Source: Holder and Lewis, 2003)

5.2 Results

5.2.1 V-PPase

5.2.1.1 Database analysis

DB searches with the PnVPP protein sequence against the *P. sojae*, *P. ramorum* and *T. pseudonana* genomes (http://genome.jgi-psf.org/) revealed two homologues of the *V-PPase* gene in all species. The correct translational start site of *TpVPP1* could not be determined due to poor sequence reads at the 5' end. Searches for *P. infestans* EST contigs using VMD (http://phytophthora.vbi.vt.edu/) revealed three contigs of varying length. Two contigs contained less than 350 aa, whereas the third one displayed more than 700 aa. However, it is possible that the contig is still incomplete considering the lengths of other V-PPases (approximately 800–1000 aa). The relevant information of the DB analysis is summarised in Table 5.2.

Table 5.2 Results of V-PPase DB analysis (*P. sojae*, *P. ramorum* and *T. pseudonana* – JGI, *P. infestans* - VMD)

			Location	
Species	**Assigned name**	**Name**	**Scaffold**	**Position**
P. sojae	estExt_fgenesh1_pm.C_90025	PsVPP1	9	744112-747201
	estExt_Genewise1.C_900043	PsVPP2	90	81416-83739
P. ramorum	fgenesh1_pm.C_scaffold_2000017	PrVPP1	2	520864-523253
	fgenesh1_pm.C_scaffold_10000002	PrVPP2	10	40960-43542
T. pseudonana	newV2.0.genewise.13.39.1	TpVPP1	13	286222-288807
	newV2.0.genewise.8.971.1	TpVPP2	8	377975-380140
P. infestans	CL1Contig3651	PiVPP1	-	-
	CL62Contig1	PiVPP2a	-	-
	CL790Contig	PiVPP2b	-	-

DNA sequences obtained were translated and basic parameters such as predicted MW and theoretical pI were computed using the ProtParam tool on the Expasy server (Table 5.3 - http://ca.expasy.org/tools/protparam.html). *PsVPP2* displayed two in-frame ATG start codons, resulting in proteins of 812 and 877 aa, respectively. However, based on codon usage and promoter element analysis (e.g. 19 bp motif), it seems likely that the start codon, resulting in the shorter 812 aa protein, is the translational start site. Introns are present in some of the sequences. *PsVPP2* and *PrVPP2* contain a single intron, whereas the *T. pseudonana* sequences contain three and four introns, respectively. A multiple sequence alignment with the two short *P. infestans* contigs revealed that they

represent the *PiVPP2* gene (data not shown). However, as contigs do not overlap and represent the complete sequence, they were not included in further analysis. DNA and protein sequence information can be found in Appendix F.

Table 5.3 Computed parameters of sequences obtained by DB analysis

Protein	Introns	Length in aa	Predicted MW in kDa	Predicted pI
PsVPP1	0	769	80.4	5.30
PsVPP2	1	812	84.7	5.18
PrVPP1	0	769	80.5	5.24
PrVPP2	1	801	84.0	5.55
TpVPP1	4	769	80.9	5.29
TpVPP2	3	798	82.8	4.48
PiVPP1	?	767	80.4	5.44

5.2.1.2 Comparison of proteins: Similarities and multiple sequence alignments

Identities and similarities between protein sequences were calculated by global pairwise alignments using the MatGAT program. Table 5.4 shows that *Phytophthora* sequences form two distinct groups (PsVPP1, PrVPP1, PiVPP1, and PnVPP, PsVPP2, PrVPP2,), sharing sequence identities of 92-96% and 93-94% respectively. TpVPP2 shows only 33-39% sequence identity to all other sequences, whereas TpVPP1 shows higher identity towards the VPP1 cluster (54-55%). To further analyse the protein sequences, a multiple sequence alignment profile was generated using the program ClustalX (Thompson *et al.*, 1997). Conserved regions were observed in all sequences, however, Figure 5.3 shows clearly the clustering of sequences into two groups and the two sequence divergent *T. pseudonana* proteins. Figure 5.4 and Figure 5.5 show alignments of the two *Phytophthora* groups only, emphasising the high homology between the group members.

Table 5.4 Identities and similarities between V-PPase proteins obtained by MatGat 2.01

			Identity in %							
			1	2	3	4	5	6	7	8
Similarity in %	1	PnVPP		33	94	34	93	33	34	33
	2	PsVPP1	52		33	92	33	92	55	38
	3	PsVPP2	96	52		33	93	32	34	33
	4	PrVPP1	52	94	51		33	96	55	39
	5	PrVPP2	97	52	96	52		33	35	33
	6	PiVPP	52	94	50	98	52		54	38
	7	TpVPP1	52	69	52	71	52	71		36
	8	TpVPP2	52	56	52	56	52	56	55	

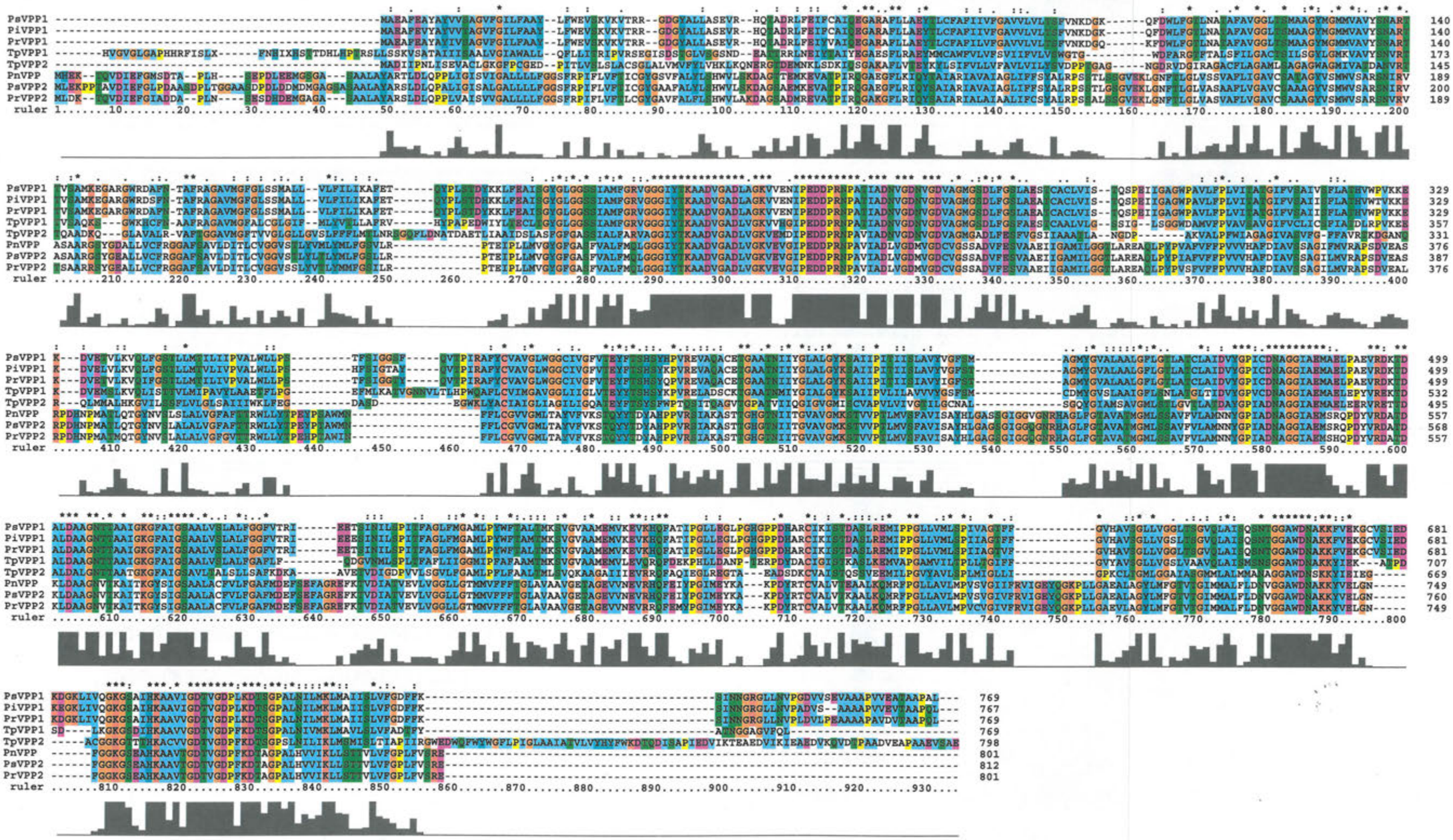


Figure 5.3 Multiple sequence alignment of *P. nicotianae*, *P. sojae*, *P. ramorum*, *P. infestans* and *T. pseudonana* V-PPase protein sequences generated by ClustalX. The degree of conservation in each column is denoted above the alignment. The asterisk [*] indicates positions that have fully conserved residues in all sequences, whereas [:] and [.] represent conservative and semi-conservative substitutions, respectively (for details see http://bips.u-strasbg.fr/fr/Documentation/ClustalX/).

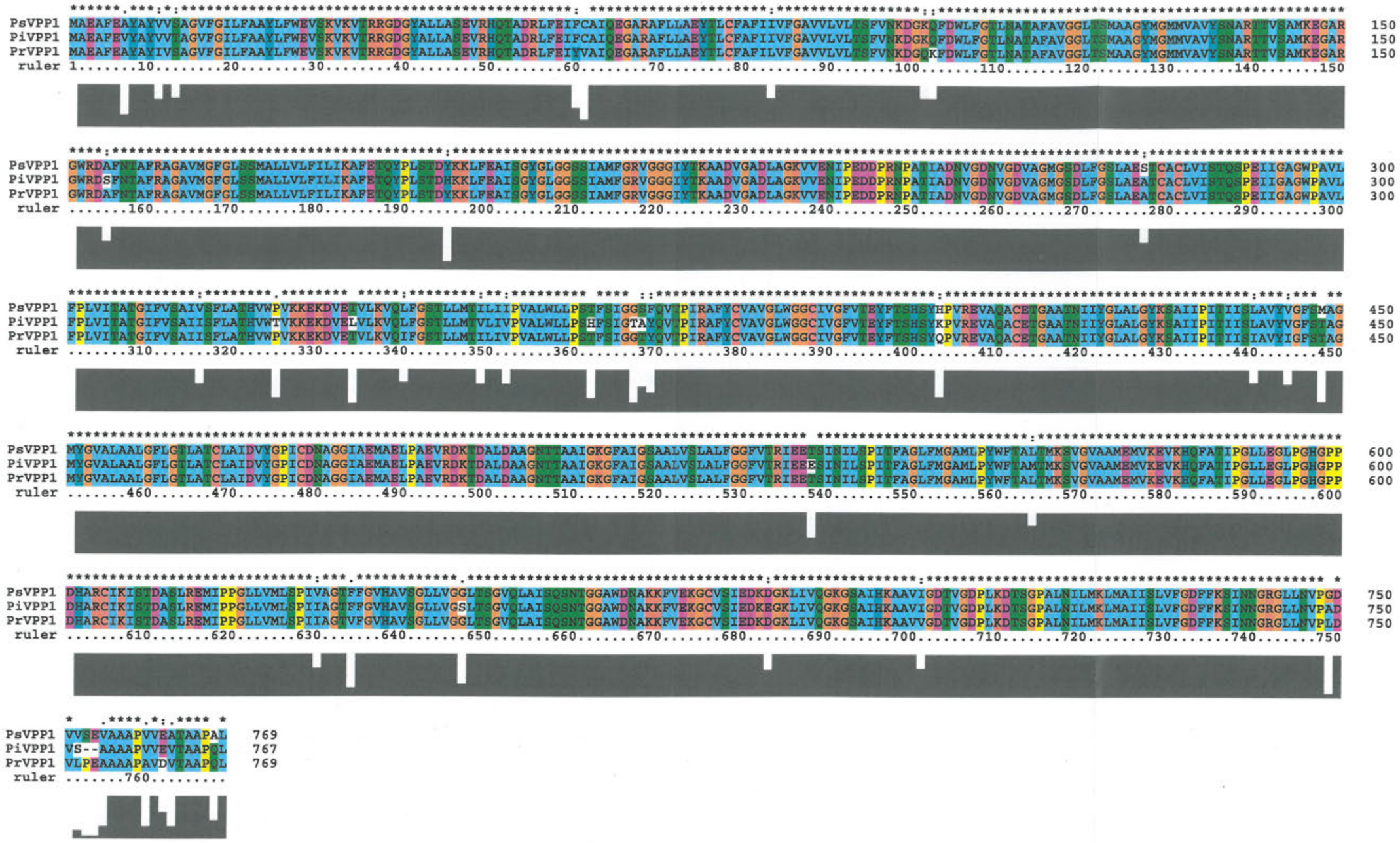


Figure 5.4 Multiple sequence alignment of *P. infestans*, *P. sojae* and *P. ramorum* VPP1 group protein sequences generated by ClustalX. The degree of conservation in each column is denoted above the alignment. The asterisk [*] indicates positions that have fully conserved residues in all sequences, whereas [:] and [.] represent conservative and semi-conservative substitutions, respectively.

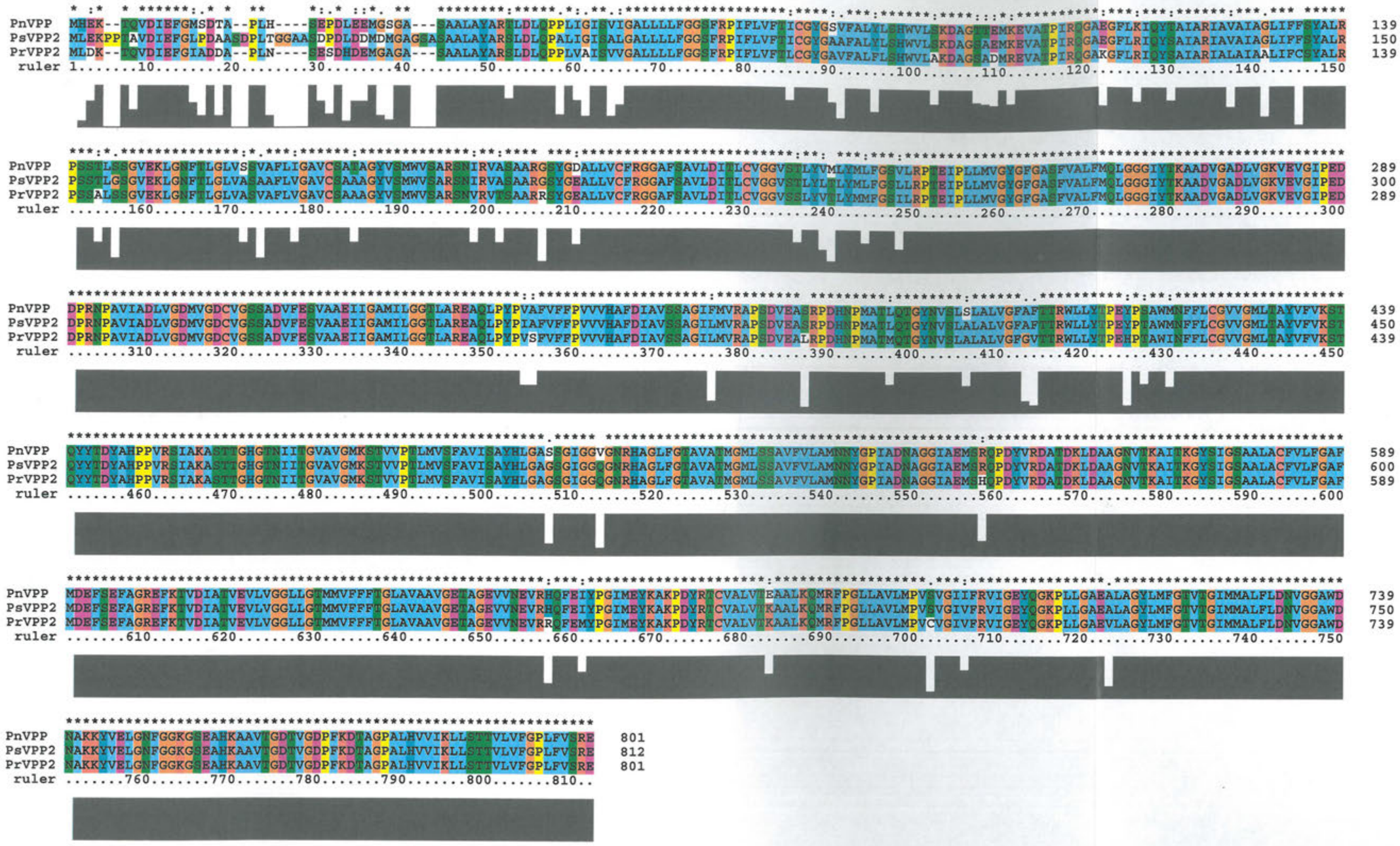


Figure 5.5 Multiple sequence alignment of *P. nicotianae*, *P. sojae* and *P. ramorum* VPP2 group protein sequences generated by ClustalX. The degree of conservation in each column is denoted above the alignment. The asterisk [*] indicates positions that have fully conserved residues in all sequences, whereas [:] and [.] represent conservative and semi-conservative substitutions, respectively.

5.2.1.3 Phylogenetic analysis

In order to represent a variety of organisms in the phylogenetic analysis the NCBI GenBank DB was mined for V-PPase sequences. A total of 29 homologous protein sequences from plants, algae, proteobacteria and archaea (see Appendix F.9) together with the six *Phytophthora* and two *T. pseudonana* sequences were used to construct a phylogenetic tree. A multiple sequence alignment was generated using ClustalX and gaps in the alignment were trimmed (i.e. only columns with a density of >50% were retained) using the program parseClustal. The resulting alignment was used to generate a NJ phylogram with 1000 bootstrap replicates (Figure 5.6). The three *Phytophthora* sequences of the VPP2 group are clustered together and appear most closely related to *P. falciparum* PfVP2 and *A. thaliana* AVP2 and AVP3. This cluster is furthermore branched off the archaeal and proteobacterial V-PPases (*P. aerophilum*, *R. rubrum* = *S. coelicolor*), strengthening the idea that the PnVPP protein analysed in this study belongs to the Type II V-PPase cluster. Members of the second *Phytophthora* cluster, VPP1, are grouped together and fall into the Type I V-PPase cluster. They appear to be most homologous to the diatom *T. pseudonana* TpVPP1 protein. The whole VPPs/TpVPP1 cluster in turn branches off between the *P. falciparum* PfVP1/*T. gondii* and *T. brucei*/*T. cruzi*/*Leishmania major* clusters. Interestingly, TpVPP2, the second *T. pseudonana* V-PPase protein, is clustered with the algal V-PPase from *A. mediterranea* and the bacterial V-PPase from *C. hydrogenoformans*, and not with the *Phytophthora* VPP2 group.

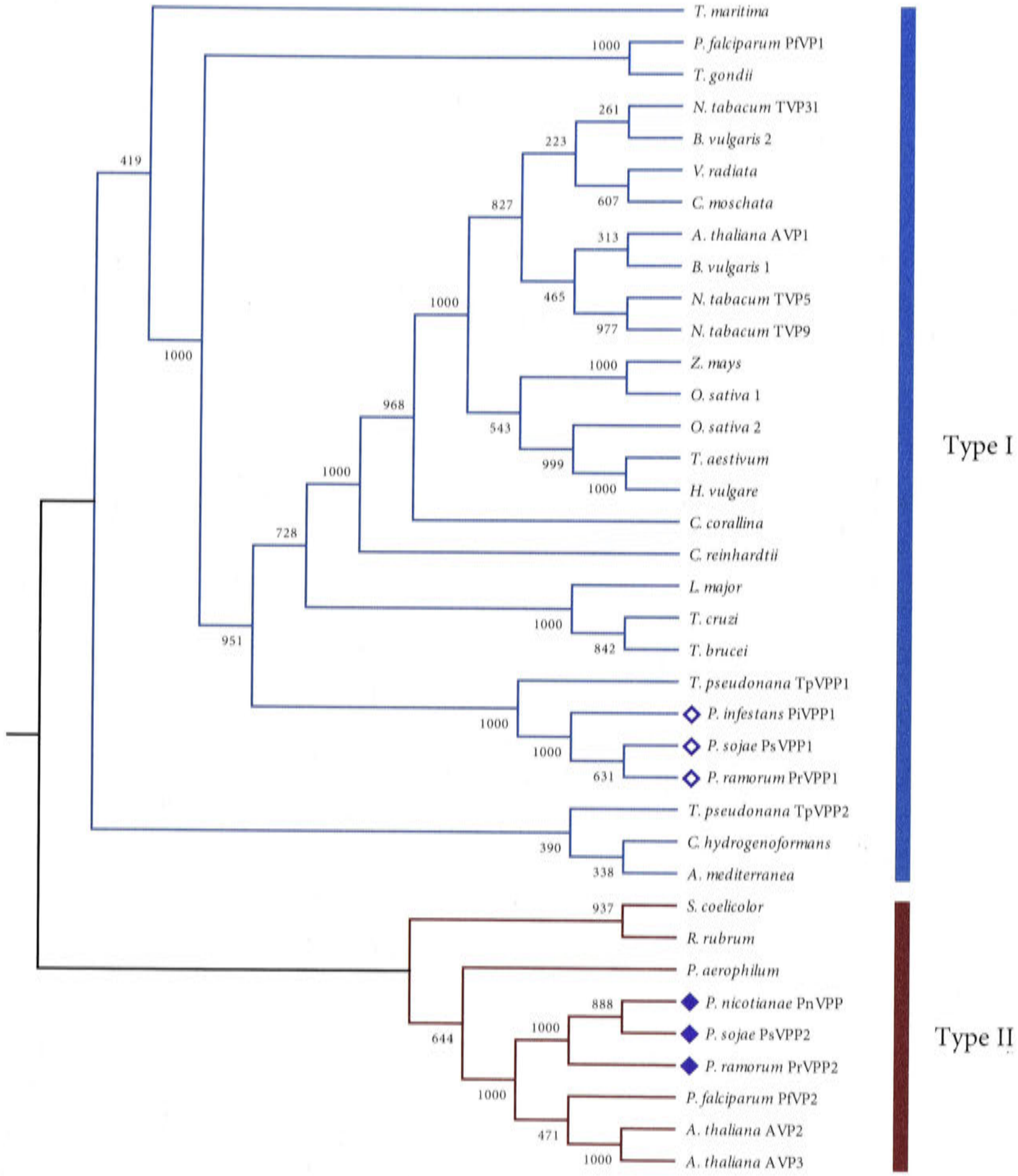


Figure 5.6 Phylogenetic comparison of *P. nicotianae*, *P. sojae*, *P. ramorum*, *P. infestans*, *T. pseudonana* and other V-PPases. V-PPase sequences were aligned using ClustalX. The alignment was trimmed using the program parseClustal (courtesy of Allan Dickerman, VBI, Blacksburg, VA, USA) and was subjected to phylogenetic analysis by the NJ distance method (PHYLIP, Felsenstein, 1989). Analysis was subjected to 1000 bootstrap replicates. Sequences used for phylogenetic tree construction are listed in Appendix F.9. *Phytophthora* sequences of the VPP1 group are indicated by an open diamond, whereas VPP2 group members are denoted by a filled diamond.

5.2.2 Aquaporins

5.2.2.1 Database analysis

JGI DB analysis of the two *Phytophthora* genomes using keywords 'aquaporin' and 'MIP' as well as searches with the PnAq2 protein sequence revealed 35 and 36 homologous sequences in the *P. sojae* and *P. ramorum* genome, respectively. Figure 5.7 shows an example of the JGI genome browser representing a *P. sojae* genomic region that contains four aquaporin genes. Predicted gene models of all genes were thoroughly examined and either verified or adjusted by analysis of the predicted transcript and the translated protein sequence using the program BioEdit (Hall, 1999). In the case of ambiguous, in-frame ATG start codons, the regions around the potential translational start sites were examined for promoter elements such as the *Phytophthora* specific 19 bp motif, TATA-box, CAAT-box and CT-rich regions. Furthermore, the GC contents of sequence stretches were taken into consideration when deciding on the most likely translational start site. A number of sequences were excluded from further analysis due to unresolvable sequence issues (e.g. apparent wrong scaffold assembly, falsely predicted introns). Furthermore, some duplicates displaying identical 3' and 5' UTRs and gene sequences were excluded. Due to this elimination, only 23 out of 35 and 22 out of 36 predicted genes of *P. sojae* and *P. ramorum*, respectively, were included in later analysis (Table 5.5 and Table 5.6). The search of the *T. pseudonana* genome resulted in one homologous sequence only (Table 5.7). However, due to two very different predicted gene models (e.g. very short vs. very long sequence, no intron vs. long intron), and a read through into another gene without an apparent stop codon, it was excluded from further analysis.

Mining for *P. infestans* aquaporin EST contigs using VMD revealed 17 non-redundant contigs of varying lengths (Table 5.8, 3 < 100 aa, 9 < 200 aa, 4 < 300, 1 > 300). Two contigs (PiAq8, PiAq18) were selected for further analyses as they displayed sufficient sequence length.

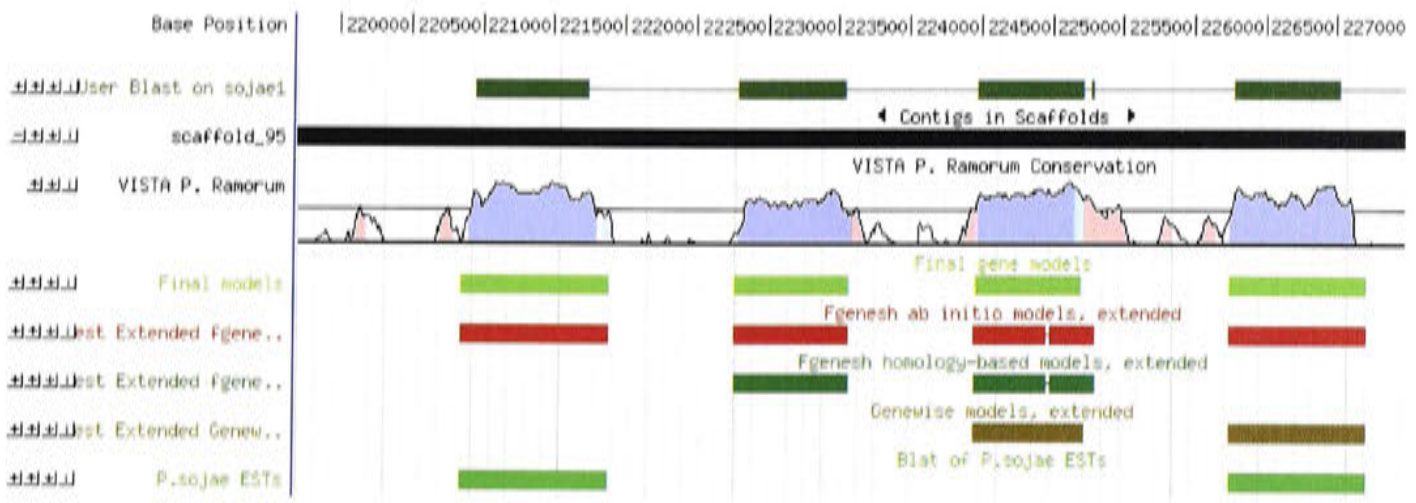


Figure 5.7 JGI genome browser representing a *P. sojae* genomic region with four aquaporin genes. Shown is the conservation of the region in *P. ramorum*, various gene models and available EST sequences.

Table 5.5 Aquaporin sequences of *P. sojae* obtained by DB analysis (JGI)

P. sojae				
Scaffold	**Name**	**Assigned Name**	**Position**	**Length**
1	**PsAq23**	estExt_fgenesh1_pm.C_10016	1091953-1093503	1667745
	PsAq11	estExt_fgenesh1_pm.C_10018	1111419-1112942	
10	**PsAq14**	estExt_fgenesh1_pm.C_100023	900048-901484	1027410
13	**PsAq15**	estExt_fgenesh1_pm.C_130007	389757-391138	940594
	PsAq2	estExt_Genewise1.C_130061	931456-932592	
18	**PsAq7**	estExt_fgenesh1_pm.C_180005	339647-340560	893458
86	**PsAq27**	estExt_fgenesh1_pg.C_860075	266332-267462	323535
	PsAq4	estExt_fgenesh1_pm.C_860008	270230-271374	
	PsAq9	estExt_fgenesh1_pg.C_860081	282642-283956	
	PsAq30	estExt_fgenesh1_pg.C_860082	285146-286201	
	PsAq22	estExt_fgenesh1_pg.C_860085	288150-290291	
	PsAq16	estExt_fgenesh1_pm.C_860010	320343-321432	
95	**PsAq17**	estExt_fgenesh1_pg.C_950046	220246-221503	330242
	PsAq13	estExt_fgenesh1_pm.C_950004	222185-223164	
	PsAq1	estExt_Genewise1.C_950130	223678-225021	
	PsAq18	estExt_fgenesh1_pg.C_950049	225355-227103	
	PsAq10	estExt_fgenesh1_pg.C_950050	229414-230544	
	PsAq21	estExt_fgenesh1_pm.C_950010	297822-299159	
	PsAq32	estExt_fgenesh1_pg.C_950074	299377-300201	
	PsAq29	estExt_fgenesh1_pg.C_950075	301234-303068	
418	**PsAq12**	estExt_fgenesh1_pm.C_4180001	2283-3412	11226
476	**PsAq6**	estExt_fgenesh1_pg.C_4760002	2732-4480	9783
	PsAq5	estExt_fgenesh1_pm.C_4760001	5128-6394	

Table 5.6 Aquaporin sequences of *P. ramorum* obtained by DB analysis (JGI)

P. ramorum				
Scaffold	**Name**	**Assigned Name**	**Position**	**Length**
6	**PrAq23**	fgenesh1_pm.C_scaffold_6000003	392012-392961	810017
29	**PrAq22**	fgenesh1_pm.C_scaffold_29000003	82665-83810	529199
46	**PrAq4**	fgenesh1_pg.C_scaffold_46000075	326281-327437	394271
	PrAq17	fgenesh1_pm.C_scaffold_46000010	358048-359309	
	PrAq21	fgenesh1_pg.C_scaffold_46000083	359513-360705	
	PrAq13	fgenesh1_pg.C_scaffold_46000085	362703-363914	
	PrAq7	fgenesh1_pm.C_scaffold_46000011	365730-366741	
	PrAq16	fgenesh1_pm.C_scaffold_46000012	367616-369111	
	PrAq18	fgenesh1_pm.C_scaffold_46000013	370343-371637	
102	**PrAq6**	fgenesh1_pm.C_scaffold_102000001	33766-34679	162604
	PrAq15	fgenesh1_pm.C_scaffold_102000002	39610-40980	
114	**PrAq11**	fgenesh1_pm.C_scaffold_114000002	42008-43645	137664
	PrAq2	fgenesh1_pm.C_scaffold_114000003	43706-44971	
	PrAq8	fgenesh1_pg.C_scaffold_114000013	45228-46805	
	PrAq24	fgenesh1_pg.C_scaffold_114000014	47278-48344	
	PrAq12	fgenesh1_pg.C_scaffold_114000015	50169-51277	
	PrAq9	fgenesh1_pm.C_scaffold_114000004	104609-105924	
	PrAq26	fgenesh1_pm.C_scaffold_114000005	106024-107020	
	PrAq25	fgenesh1_pm.C_scaffold_114000006	108430-109430	
660	**PrAq20**	fgenesh1_pm.C_scaffold_660000001	6455-7466	9249
716	**PrAq5**	fgenesh1_pm.C_scaffold_716000001	653-2013	8970
2368	**PrAq19**	fgenesh1_pg.C_scaffold_2368000002	1151-2239	3339

Table 5.7 Aquaporin sequence of *T. pseudonana* obtained by DB analysis (JGI)

			Location	
Species	Assigned name	Name	Scaffold	Position
T. pseudonana	newV2.0.grail.77.11.1	TpAq	77	30420-33065

Table 5.8 Aquaporin EST contigs of *P. infestans* obtained by DB analysis (VMD). Contigs in dark blue were selected for further analysis.

P. infestans			
Name	**Assigned Name**	**Library**	**Contig**
PiAq1	Pi_001_41068	mycelium, starved in water	-
PiAq2	Pi_001_41406	mycelium, starved in water	CL1Contig3510
PiAq3	Pi_002_42333	zoospores, purified	CL2129Contig1
PiAq4	Pi_002_42504	mycelium, starved in water	CL2366Contig1
PiAq5	Pi_002_43067	mycelium, starved in water	CL2144Contig1
PiAq6	Pi_004_46541	mycelium, non-sporulating growth	-
PiAq7	Pi_006_46892	mycelium, starved in water zoospores, purified	CL1Contig5300
PiAq8	Pi_002_47830	mycelium, starved in water	CL1Contig5332
PiAq9	Pi_004_48103	mycelium, starved in water cysts, germinating	CL1Contig5300
PiAq10	Pi_001_51528	sporangia, cleaving	CL51Contig1
PiAq12	Pi_001_60558	mycelium, infection mimic	-
PiAq13	Pi_001_61932	zoospores, purified	-
PiAq15	Pi_001_66729	sporangia, germinating	-
PiAq16	Pi_001_68038	mycelium, starved in water	CL1Contig5332
PiAq17	Pi_001_70980	mycelium, Plich medium	-
PiAq18	Pi_001_76243	sporangia, cleaving	CL51Contig1
PiAq19	Pi_001_76820	zoospores, purified	CL713Contig1

5.2.2.2 Protein characteristics

All DNA sequences retained for further analyses were translated and basic parameters such as predicted molecular weight and theoretical pI were computed using ProtParam (Table 5.9). None of the sequences contained introns. *P. sojae* protein sequences range from 274 to 320 aa in length with predicted molecular weights of 29.6 to 35.2 kDa. The majority of proteins show a predicted pI between 5 and 7, whereas four sequences are above 8. In *P. ramorum*, proteins range from 274 to 352 aa and show predicted MWs of 29.7 to 38.6 kDa. The computed pIs range from 5 to 8.15, with most proteins displaying their pI between 5 and 7. The two *P. infestans* proteins display 274 to 304 aa, a molecular weight of 29.5to33.3 kDa and a pI between 5.4 to 6.5.

DNA and protein sequence information of all genes retained for analysis can be found in Appendix G.

Table 5.9 Computed parameters of *Phytophthora* aquaporin sequences obtained by DB analysis

Protein	Predicted Length in aa	Predicted MW in kDa	Predicted pI
P. sojae			
PsAq1	290	31.5	5.18
PsAq2	299	32.7	6.51
PsAq4	318	35.2	6.20
PsAq5	299	32.5	6.98
PsAq6	312	34.1	6.20
PsAq7	304	33.2	8.65
PsAq9	304	33.3	6.14
PsAq10	311	33.7	6.15
PsAq11	302	32.6	6.24
PsAq12	312	34.1	5.84
PsAq13	302	32.6	6.04
PsAq14	298	32.7	6.79
PsAq15	317	35.0	6.20
PsAq16	313	34.2	8.17
PsAq17	300	32.2	7.18
PsAq18	294	31.5	5.77
PsAq21	274	29.6	6.50
PsAq22	320	34.2	8.13
PsAq23	299	31.8	6.43
PsAq27	316	35.2	8.13
PsAq29	274	29.7	5.14
PsAq30	311	34.2	7.61
PsAq32	274	29.8	6.53
P. infestans			
PiAq8	304	33.3	5.43
PiAq18	274	29.5	6.50
P. ramorum			
PrAq1	352	38.6	6.05
PrAq2	295	31.6	6.05
PrAq4	318	35.1	6.33
PrAq5	299	32.7	5.92
PrAq6	301	32.2	6.49
PrAq7	311	34.3	6.98
PrAq8	289	31.5	5.06
PrAq9	274	29.7	6.58
PrAq11	300	32.3	6.97
PrAq12	305	32.7	6.14
PrAq13	333	36.4	6.21
PrAq15	302	32.8	7.21
PrAq16	309	33.6	5.75
PrAq17	315	35.4	5.80
PrAq18	320	34.3	6.40
PrAq19	299	32.7	6.93
PrAq20	313	34.2	6.50
PrAq21	328	35.9	8.15
PrAq22	315	34.8	6.35
PrAq23	297	32.5	6.82
PrAq24	318	34.1	6.50
PrAq25	275	29.8	5.33
PrAq26	274	29.9	5.94

5.2.2.3 Multiple sequence alignments and conserved domains

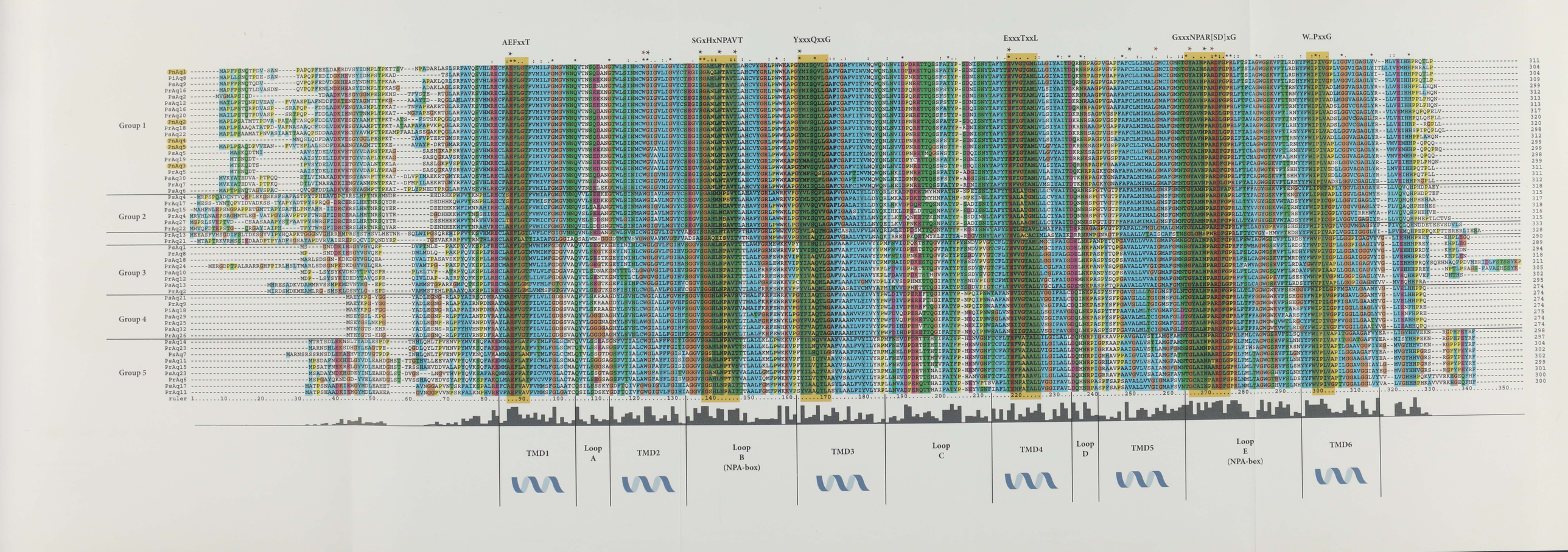
A multiple sequence alignment was generated using the program ClustalX to analyse regions of homology between all 52 *Phytophthora* aquaporin protein sequences. The alignment revealed regions of conserved aquaporin domains in all proteins. Based on sequence similarities, proteins were clustered into five groups (Figure 5.8). Sequences PrAq13 and PrAq21 were not included into any group due to high sequence divergence. The grouping of proteins and exclusion of the two *P. ramorum* sequences was confirmed by construction of a phylogenetic tree. The multiple sequence alignment was subjected to trimming by the program parseClustal and the resulting alignment was used to generate a NJ phylogram with 1000 bootstrap replicates (Figure 5.9). All five PnAq proteins analysed in this study belong to Group 1.

Conserved domains, as described in detail in Chapter 4 (4.2.3) are highlighted in the multiple sequence alignment in Figure 5.8. Changes in the motifs are summarised in Table 5.10. Members of Group 4 show the lowest overall sequence divergence in conserved motifs, closely followed by Group 3. Group 1 displays a few residue substitutions which are mostly persistent throughout the group, whereas Groups 2 and 5 seem to contain the most divergent proteins. Residue changes displayed in the column PrAq13/PrAq21 are mainly contributed by PrAq21, rendering it the most divergent sequence of all *Phytophthora* proteins.

Residues which form the selectivity filter of MIPs were inferred from the sequence alignment and are shown in

Table 5.11. As the majority of residues are conserved and match the *E. coli* GlpF residues, it was suggested that most *Phytophthora* aquaporin proteins belong to the GLP cluster and are likely to transfer glycerol and possibly also other small solutes and water.

Figure 5.8 Multiple sequence alignment of all 52 *Phytophthora* aquaporin protein sequences generated by ClustalX. The degree of conservation in each column is denoted above the alignment. The small asterisk [*] indicates positions that have fully conserved residues in all sequences, whereas [:] and [.] represent conservative and semi-conservative substitutions, respectively. Red (big) asterisks denote amino acid residues involved in formation of the narrowest constriction, whereas black (big) asterisks indicate residues that play a role in stabilisation of the pore by side-chain interaction. *Phytophthora* proteins are clustered into five groups as indicated. Orange boxes show conserved domains. Predicted TMD and loops A-E are shown below the alignment.


AEFxxT
SGxHxNPAVT
YxxxQxxG
ExxxTxxL
GxxxNPAR[SD]xG
W..PxxG
Group 1
Group 2
Group 3
Group 4
Group 5
TMD1
Loop A
TMD2
Loop B (NPA-box)
TMD3
Loop C
TMD4
Loop D
TMD5
Loop E (NPA-box)
TMD6

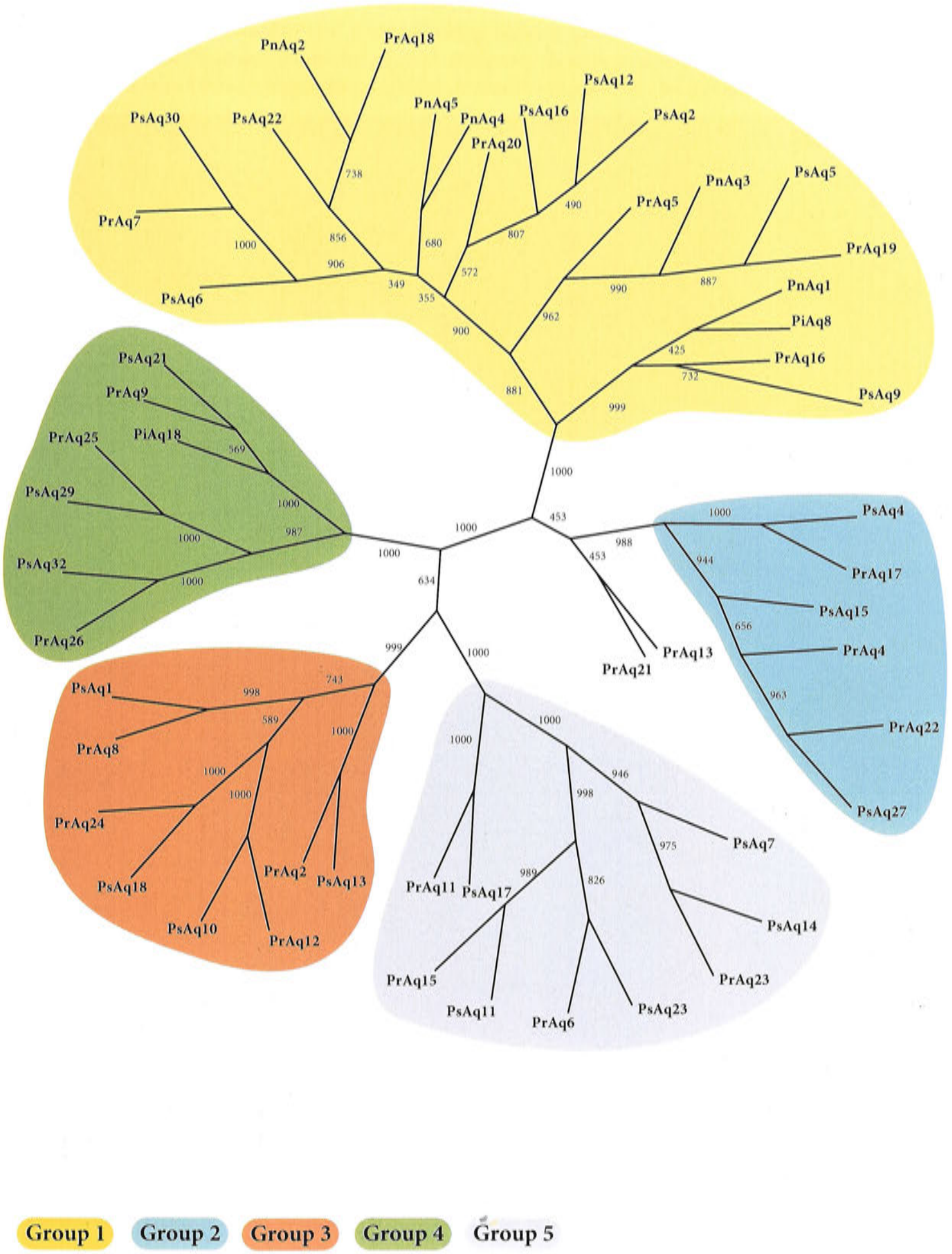


Figure 5.9 Phylogenetic comparison of 52 *Phytophthora* aquaporin proteins. Aquaporin sequences were aligned using ClustalX. The alignment was trimmed using the program parseClustal and was subjected to phylogenetic analysis by the NJ distance method (PHYLIP, Felsenstein, 1989). Analysis was subjected to 1000 bootstrap replicates.

Table 5.10 Conserved domains and corresponding residue substitutions in *Phytophthora* aquaporin protein sequences. Numbers in brackets below the group denomination indicate the number of group members. Numbers following a substitute residue denote the number of changes within the group.

Conserved motif		Group 1 (20)	Group 2 (6)	Group 3 (8)	Group 4 (7)	Group 5 (9)	PrAq13 PrAq21
AEFxxT (TMD1)	A		S-2 T-1				
	E	–		–	–		–
	F						
	T					M-1 V-2	
G (TMD2)	G	–	–	–	–	–	–
YxxxQxxG (TMD3)	Y						
	Q	–	–		–		–
	G			A-1		A-8	
ExxxTxxL (TMD4)	E						
	T	–	A-1	–	–		C-1
	L					A-4	I-1
G (TMD5)	G	C-15 A-5	A-6	A-8	A-4	A-2	A-2
W…PxxG (TMD6)	W		L-2				
	P				–		
	G	A-14 S-2	A-2	A-3		A-8	A-1
SGxHxNPAVT (Loop B)	S						
	G						
	H	Q-1 N-16				N-1	Q-1
	N						I-1
	P	T -20	C-2 S-2				T-1 S-1
	A		S-2			T-4	V-1
	V			I-2		I-4	I-1
	T				S-3	S-1	
GxxxNPAR(S/D)xG (Loop E)	G						
	N						S-1
	P					A-4	
	A					T-1	N-1
	R					K-1	
	S/D	(D)	(D)	(D)	(D)	(D)	(D)
	G		A-2				

Table 5.11 Residues forming the selectivity filter of *Phytophthora* aquaporin proteins

MIP	Organism	Reference	Position ①	②	③	④
GlpF	***E. coli***	Fu *et al.*, 2000	**W48**	**G191**	**R206**	**F200**
PnAq2	***P. nicotianae***	-	**W109**	**G243**	**R258**	**Y252**
Group1	*Phytophthora*		W	G	R	Y
Group2	*Phytophthora*		W	G	R	Y
Group3	*Phytophthora*		W	G	R	Y/F
Group4	*Phytophthora*		W	G	R/K	Y
Group5	*Phytophthora*		W	G/A	R	L/C

5.2.2.4 Comparison of proteins: Identities and similarities

Identities and similarities between protein sequences were calculated by global pairwise alignments using the MatGAT program.

Table 5.12 shows members of each group and protein similarities and identities within each group. All five groups show generally high identities and similarities (~70-90%) between group members. Interestingly, even though proteins PrAq13 and PrAq21 were outgrouped and clustered together in the phylogenetic analysis, they share only 40% sequence identity and 59% similarity. In Table 5.13, members of groups between species were compared. High homologies were observed between proteins of corresponding groups (identities up to 98%), whereas identities between members of different groups dropped significantly to around 30-50%.

Table 5.12 Similarities/Identities of *Phytophthora* aquaporin protein sequences

Organism	Group 1	2	3	4	5
P. *nicotianae*	PnAq1 PnAq2 PnAq3 PnAq4 PnAq5	–	–	–	–
Id	71-84%				
Sim	81-88%				
P. sojae	PsAq2 PsAq 5 PsAq 6 PsAq 9 PsAq12 PsAq16 PsAq22 PsAq30	PsAq4 PsAq15 PsAq27	PsAq1 PsAq10 PsAq13 PsAq18	PsAq21 PsAq29 PsAq32	PsAq7 PsAq11 PsAq14 PsAq17 PsAq23
Id	69-90%	67-75%	56-81%	72-83%	51-75%
Sim	79-91%	77-84%	72-87%	83-90%	64-86%
P. ramorum	PrAq5 PrAq7 PrAq16 PrAq18 PrAq19 PrAq20	PrAq4 PrAq17 PrAq22	PrAq2 PrAq8 PrAq12 PrAq24	PrAq9 PrAq25 PrAq26	PrAq6 PrAq11 PrAq15 PrAq23
Id	68-89%	68-78%	58-73%	74-87%	48-76%
Sim	80-94%	80-87%	74-83%	85-92%	64-88%

Table 5.13 Comparison of *P. sojae* and *P. ramorum* protein sequences

		P. ramorum 1	2	3	4	5	
P. sojae	1	**68-93%**	52-60%	40-51%	39-45%	35-44%	Id
		79-96%	66-74%	58-68%	56-63%	54-63%	Sim
	2	52-60%	**67-81%**	38-44%	37-40%	33-40%	Id
		65-72%	**79-90%**	55-60%	51-56%	50-57%	Sim
	3	39-50%	36-43%	**56-94%**	55-47%	41-51%	Id
		60-67%	57-61%	**72-98%**	64-73%	58-68%	Sim
	4	40-45%	37-41%	48-55%	**72-96%**	37-48%	Id
		55-63%	55-57%	63-73%	**83-97%**	54-63%	Sim
	5	35-43%	32-38%	42-49%	39-48%	**48-91%**	Id
		52-60%	52-56%	58-67%	56-62%	**64-96%**	Sim

Comparison of *P. sojae* and *P. ramorum* Group 1 proteins also includes PnAq sequences

5.2.2.5 Southern blot analysis

In order to ascertain whether the *P. nicotianae* genome also contains aquaporin genes encoding for proteins of other groups, Southern blot analysis with probes derived from *P. sojae* DNA sequences was carried out. Individual DNA multiple sequence alignments of Groups 2-5 were generated using ClustalX. Alignments were analysed for homologous regions within each group, displaying as little homology to other groups as possible. Probes were PCR amplified from *P. sojae* genomic DNA and equal amounts of probes were labelled with DIG. To represent Group 1 in Southern blot analysis, an equal mix of *P. nicotianae* Aq1-Aq5 PCR products obtained from the SSCP analysis (see 4.2.5) was used. Labelling efficiency of all five DIG-labelled probes was tested to ensure the use of equal amounts of probes in Southern blot analysis (Figure 5.10). Two blots with genomic DNA of *P. nicotianae* and *P. sojae* digested with *Sac*I, *Sal*I and *Pst*I were prepared (Figure 5.11). Blots were cut into five segments, each containing DNA digested with all three enzymes, and individual strips were subjected to Southern blot analysis with one of the five probes (Figure 5.12). All probes derived from *P. sojae* genomic DNA showed cross-reaction with *P. nicotianae* DNA and *vice versa*. Blots labelled with probes representing Groups 1 and 2 show an almost identical band pattern, indicating that probes cross-react with members of the other group. The other three probes showed labelling of different bands in each blot, suggesting that no cross-reactions between groups occurred. Table

5.14 shows a comparison of the number of group members and the number of bands obtained by Southern blot analysis.

Results strongly indicate that the *P. nicotianae* genome contains not only the five aquaporin genes analysed in this study, but also a number of other genes representing the remaining four groups.

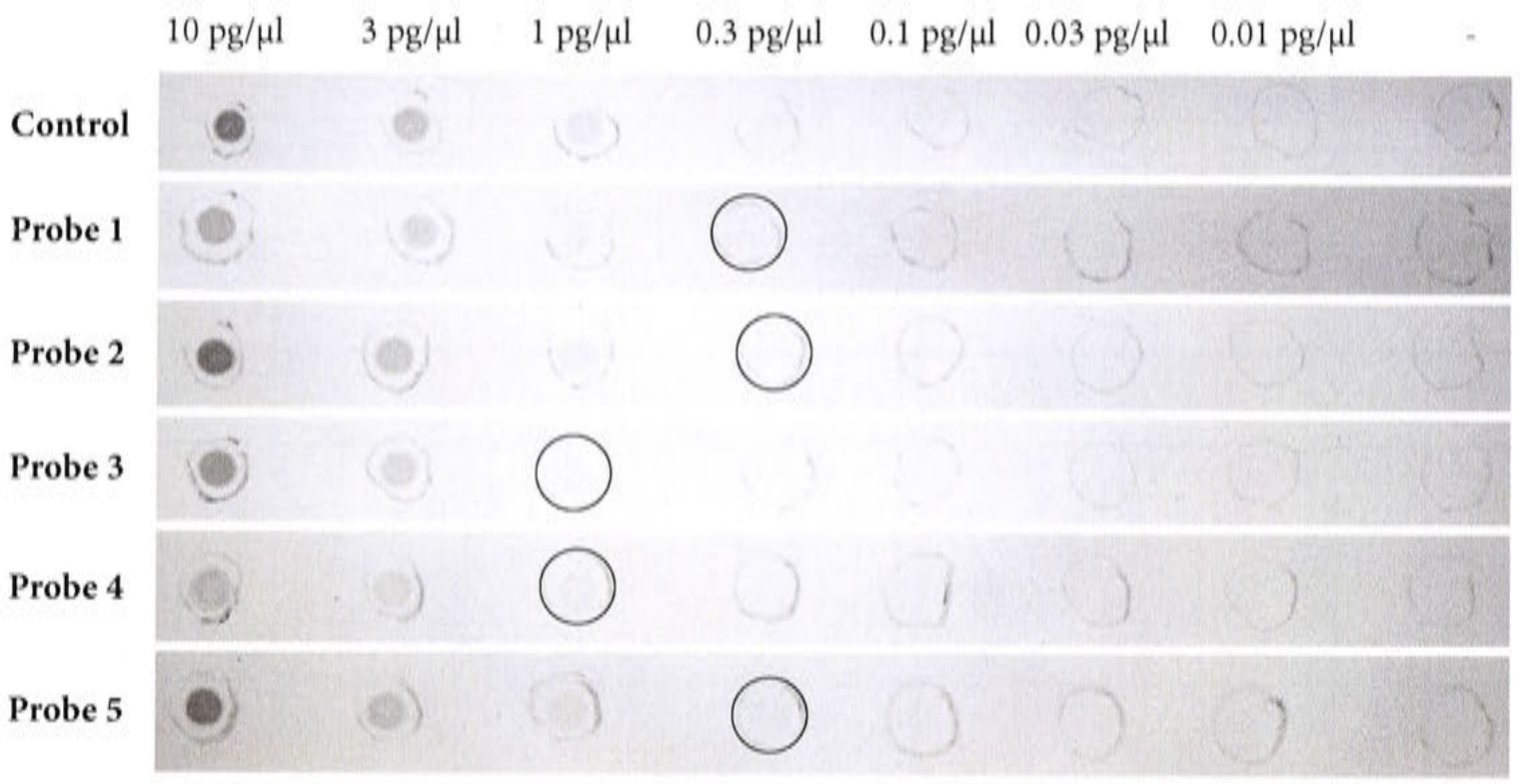


Figure 5.10 Test of labelling efficiency of DIG-labelled probes.

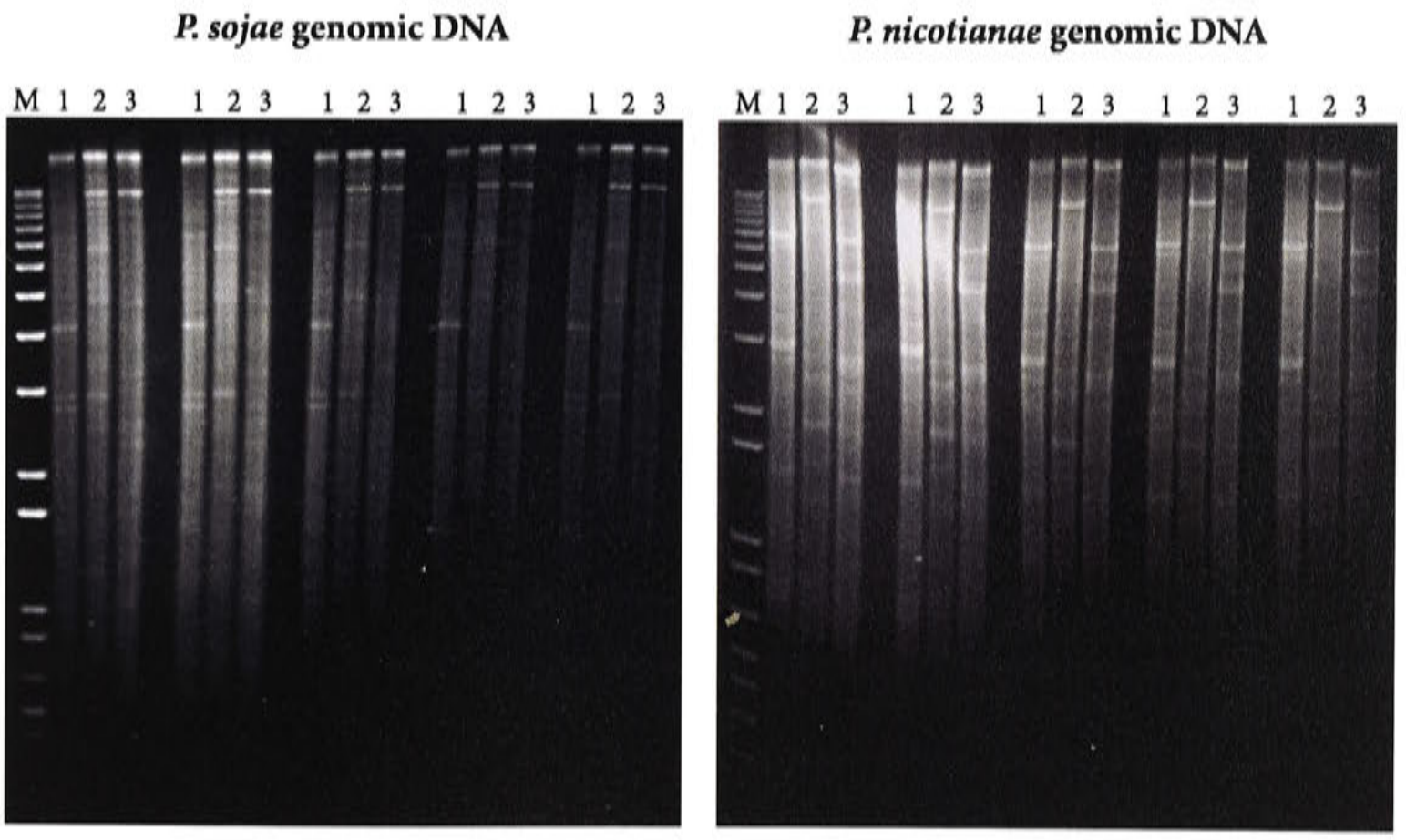


Figure 5.11 Restriction digest of *P. sojae* and *P. nicotianae* genomic DNA with enzymes *Sac*I (1), *Sal*I (2) and *Pst*I (3) for Southern blot analysis with group-specific probes. Pictures show the EtBr stained gel before transfer. M – molecular weight marker

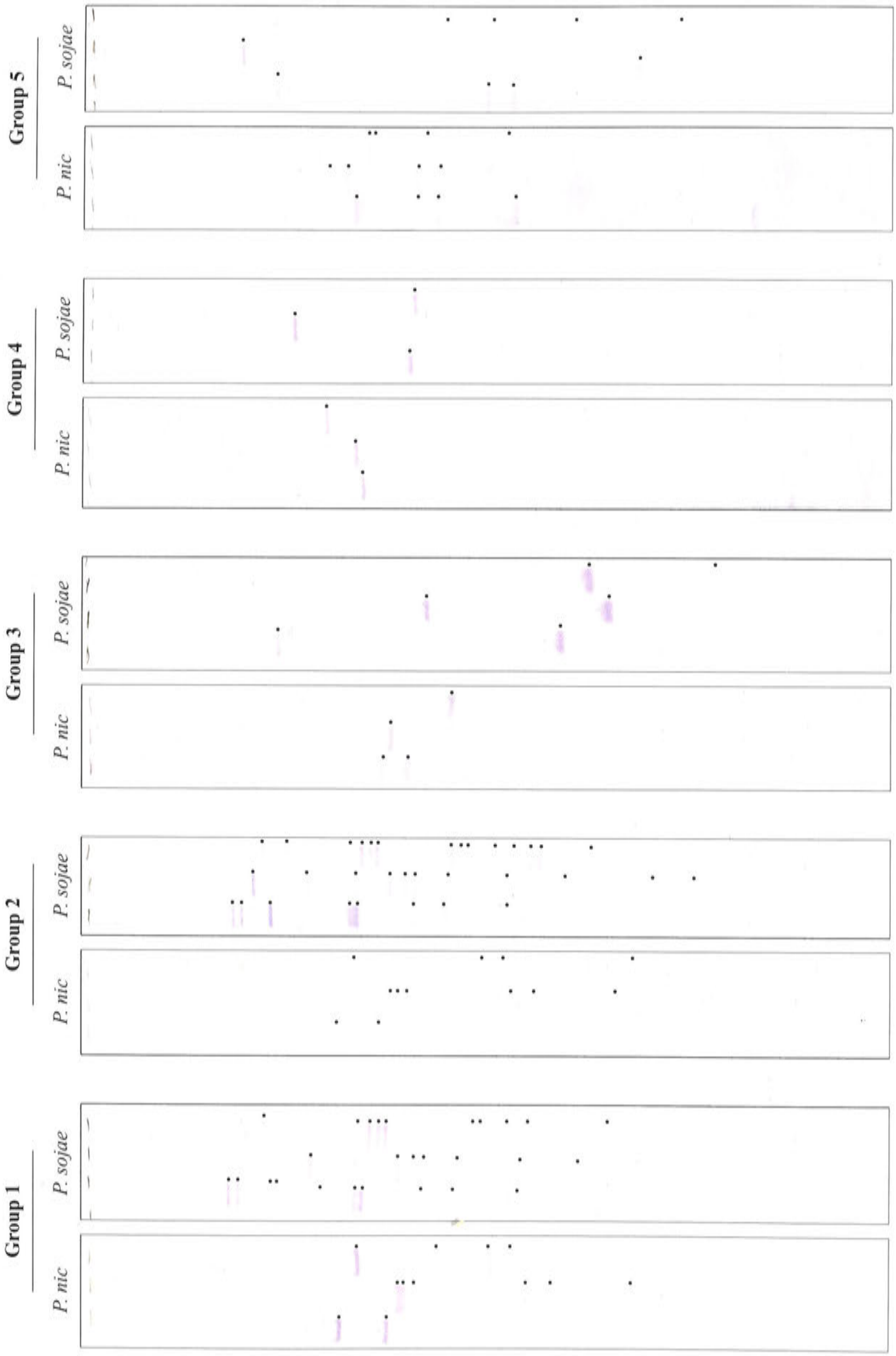


Figure 5.12 Southern blot analysis of *P. nicotianae* and *P. sojae* genomic DNA with probes representing individual groups. DNA was digested with *Sac*I (lane 1), *Sal*I (lane 2) and *Pst*I (lane 3). Dots indicate labelled bands.

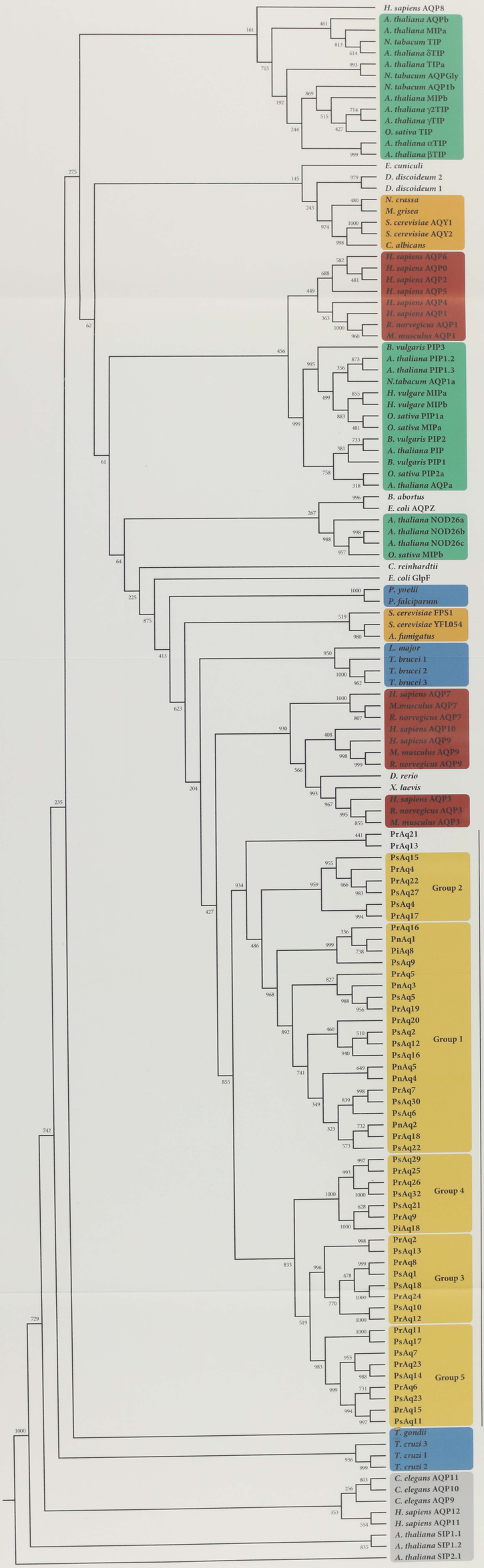
Table 5.14 Comparison of the number of group members and the numbers of bands obtained by Southern blot analysis with group-specific probes

		P. sojae				*P. nicotianae*			
		Group members #	Southern blot			Group members #	Southern blot		
			*Sac*I	*Sal*I	*Pst*I		*Sac*I	*Sal*I	*Pst*I
Probe	1	8	10	7	10	5	2	6	4
	2	3	8	11	14	–	2	6	3
	3	4	2	2	2	–	2	1	2
	4	3	1	1	1	–	1	1	1
	5	5	3	2	4	–	4	4	4

5.2.2.6 Phylogenetic analysis

To represent a variety of organisms in the phylogenetic analysis, the NCBI GenBank DB was mined for aquaporin sequences. A total of 84 homologous sequences from mammals, plants, algae, fungi, apicomplexans, kinetoplastida and bacteria (see Appendix G.8) together with the 52 *Phytophthora* sequences were used to construct a phylogenetic tree. A multiple sequence alignment was generated using ClustalX and gaps in the alignment were trimmed (i.e. only columns with a density of >50% were retained) using the program parseClustal. The resulting alignment was used to generate a NJ phylogram with 1000 bootstrap replicates (Figure 5.13). All 52 *Phytophthora* sequences are grouped together and appear most closely related to the clade comprised of AQP3, AQP7, AQP9 and AQP10 sequences. The five subgroups contain mostly orthologues (i.e. the same gene in different species, which evolved from the same ancestral gene). Examples are proteins PrAq16, PnAq1, PiAq8 and PsAq9 in Group 1, and PsAq21, PrAq9 and PiAq18 in Group 4. Paralogues (i.e. two homologues genes in the same species obtained by gene duplication), were also observed in some groups, for example PnAq4, PnAq5 and PsAq2, PsAq12 in Group 1. The two *P. ramorum* proteins that were excluded from the initial grouping do not have any apparent orthologues in *P. sojae* and appear to be the most distant members of the clade.

Figure 5.13 Phylogenetic comparison of *P. nicotianae*, *P. sojae*, *P. ramorum*, *P. infestans* and other aquaporins. Aquaporin sequences were aligned using ClustalX. The alignment was trimmed using the program parseClustal and was subjected to phylogenetic analysis by the NJ distance method. Analysis was subjected to 1000 bootstrap replicates. Sequences used for phylogenetic tree construction are listed in Appendix G.8. Colours represent the following organisms: green – plants, red – mammals, yellow – apicomplexans and kinetoplastida, orange fungi, grey – sub-aquaporins


H. sapiens AQP8
A. thaliana AQPb
A. thaliana MIPa
N. tabacum TIP
A. thaliana δTIP
A. thaliana TIPa
N. tabacum AQPGly
N. tabacum AQP1b
A. thaliana MIPb
A. thaliana γ2TIP
A. thaliana γTIP
O. sativa TIP
A. thaliana αTIP
A. thaliana βTIP
E. cuniculi
D. discoideum 2
D. discoideum 1
N. crassa
M. grisea
S. cerevisiae AQY1
S. cerevisiae AQY2
C. albicans
H. sapiens AQP6
H. sapiens AQP0
H. sapiens AQP2
H. sapiens AQP5
H. sapiens AQP4
H. sapiens AQP1
R. norvegicus AQP1
M. musculus AQP1
B. vulgaris PIP3
A. thaliana PIP1.2
A. thaliana PIP1.3
N.tabacum AQP1a
H. vulgare MIPa
H. vulgare MIPb
O. sativa PIP1a
O. sativa MIPa
B. vulgaris PIP2
A. thaliana PIP
B. vulgaris PIP1
O. sativa PIP2a
A. thaliana AQPa
B. abortus
E. coli AQPZ
A. thaliana NOD26a
A. thaliana NOD26b
A. thaliana NOD26c
O. sativa MIPb
C. reinhardtii
E. coli GlpF
P. yoelii
P. falciparum
S. cerevisiae FPS1
S. cerevisiae YFL054
A. fumigatus
L. major
T. brucei 1
T. brucei 2
T. brucei 3
H. sapiens AQP7
M.musculus AQP7
R. norvegicus AQP7
H. sapiens AQP10
H. sapiens AQP9
M. musculus AQP9
R. norvegicus AQP9
D. rerio
X. laevis
H. sapiens AQP3
R. norvegicus AQP3
M. musculus AQP3
PrAq21
PrAq13
PsAq15
PrAq4
PrAq22
PsAq27
PsAq4
PrAq17
Group 2
PrAq16
PnAq1
PiAq8
PsAq9
PrAq5
PnAq3
PsAq5
PrAq19
PrAq20
PsAq2
PsAq12
PsAq16
PnAq5
PnAq4
PrAq7
PsAq30
PsAq6
PnAq2
PrAq18
PsAq22
Group 1
PsAq29
PrAq25
PrAq26
PsAq32
PsAq21
PrAq9
PiAq18
Group 4
PrAq2
PsAq13
PrAq8
PsAq1
PsAq18
PrAq24
PsAq10
PrAq12
Group 3
PrAq11
PsAq17
PsAq7
PrAq23
PsAq14
PrAq6
PsAq23
PrAq15
PsAq11
Group 5
T. gondii
T. cruzi 3
T. cruzi 1
T. cruzi 2
C. elegans AQP11
C. elegans AQP10
C. elegans AQP9
H. sapiens AQP12
H. sapiens AQP11
A. thaliana SIP1.1
A. thaliana SIP1.2
A. thaliana SIP2.1

5.3 Discussion

5.3.1 V-PPase

V-PPases are divided into two groups, Type I and Type II enzymes, depending on their requirement for K^+ to function, as discussed in detail in Chapter 3 (3.1.3). Based on sequence, the *P. nicotianae* V-PPase, PnVPP, was predicted to belong to the Type II group. Further bioinformatic analysis, including data from the genomes of *P. sojae* and *P. ramorum*, revealed that the two *Phytophthora* genomes contain two divergent V-PPase sequences, representing both groups. The *P. nicotianae* protein and its orthologues, PsVPP2 and PrVPP2, are clustered together in the Type II clade, whereas the three *Phytophthora* proteins belonging to the VPP1 group (PsVPP1, PrVPP1 and PiVPP1) cluster with high confidence in the Type I clade. This situation is relatively unusual as most organisms contain only one V-PPase. Bacteria predominantly have a Type II V-PPase and apicomplexans, kinetoplastida and plants have a Type I V-PPase. However, in a few cases the occurrence of both V-PPases in a single organism has been described. Two species of *Plasmodium*, *P. falciparum* and *P. yoelii*, have been shown to contain two V-PPase genes and in *P. falciparum*, thorough characterisation of the genes has confirmed that they represent both types of V-PPase (McIntosh *et al.*, 2001; McIntosh and Vaidya, 2002). In *A. thaliana* three V-PPase sequences have been reported, one belonging to the Type I group (AVP1) and the other two, AVP2 and AVP3, falling into the Type II group (Sarafian *et al.*, 1992; Drozdowicz *et al.*, 2000; Mitsuda *et al.*, 2001). *Phytophthora* Type I proteins analysed in this study, share >92% sequence identity (>94% similarity) to each other, but only <33% identity (<52% similarity) to the *Phytophthora* Type II proteins. Thus, it is evident that the V-PPase proteins are more closely related to their orthologues than to their paralogues.

Type I V-PPase proteins of *Phytophthora* species grouped with the V-PPase, TpVPP1, of *T. pseudonana*, a marine diatom which also belongs to the kingdom Stramenopila. Surprisingly, the second V-PPase in the diatom, TpVPP2, did not cluster with the Type II *Phytophthora* proteins as expected, but with the Type I V-PPases of *A. mediterranea* and *C. hydrogenoformans*. The predicted gene model for the *T. pseudonana* VPP2 contains three introns. It is conceivable that incorrect intron prediction by the gene model prediction program led to a partially wrong protein sequence, thus resulting in a protein sequence that it is more similar to the Type I than the Type II proteins. As no EST data

are available for this *T. pseudonana* gene, the gene translation proposed here can neither be verified nor dismissed. In order to obtain relevant data, cDNA information needs to be attained by, for example cDNA library screening or RT-PCR amplification using GSPs.

Both stramenopile groups, VPP1 and VPP2, are closely related to the V-PPases in the apicomplexan species *P. falciparum*. Detailed bioinformatic analyses of, for example, ribosomal ITS sequences has greatly strengthened evidence for the close phylogenetic relationship between the Apicomplexa and Stramenopiles, a result that has received further support by recent studies of individual, protein encoding genes. The secreted *P. cinnamomi* protein PcVsv1, for example, shows high similarity to thrombospondin-type 1 repeat proteins also found in apicomplexan species such as *P. falciparum* and *Cryptosporidium parvum* (Robold and Hardham, 2005). Furthermore, a host targeting signal, showing the consensus motif RxLx (*P. falciparum*) and RxLR (*Phytophthora* species), has recently been shown to be indispensable for a large number of secretory proteins (Bhattacharjee *et al.*, 2006). By substitution of the *P. falciparum* host targeting signal sequence with that of the *Phytophthora*-specific motif, it was demonstrated that these leader sequences are interchangeable and that the *Phytophthora* motif is capable of directing efficient export of proteins from the malarial parasite cell, thus indicating a highly conserved pathogenic secretion strategy across the two kingdoms (Bhattacharjee *et al.*, 2006).

From the analysis of the V-PPases in *P. sojae* and *P. ramorum*, it seems likely that the *P. nicotianae* genome may also contain two V-PPases, one of each type. In order to elucidate whether the *P. nicotianae* genome also contains a second V-PPase gene belonging to the Type I group, one approach would be to re-screen the *P. nicotianae* genomic BAC library (Shan and Hardham, 2004) with a probe designed from the *P. sojae* and *P. ramorum* DNA sequences.

5.3.2 Aquaporins

5.3.2.1 DB analysis

The DB analysis performed in this study revealed that aquaporin genes are present in all *Phytophthora* species analysed. Searches were performed on the genomes of *P. sojae* and *P. ramorum* as well as on the EST DB of *P. infestans*. Fifty-one *Phytophthora* sequences

were retrieved, forming five groups with two sequence divergent proteins that were not included in any group. Interestingly, the genome analysis of the diatom *T. pseudonana* revealed only one aquaporin sequence. However, due to conflicting gene model predictions, the occurrence of introns and no consensus translation of the gene, the diatom sequence was excluded from further analysis. Therefore, it was not possible to analyse the relative position and evolutionary distance of the *T. pseudonana* aquaporin to the *Phytophthora* proteins.

It is conceivable that the *Phytophthora* genomes analysed contain a larger number of aquaporin genes than predicted in this study, as approximately 35% of the JGI-predicted genes were excluded from further analysis due to apparent duplications, likely scaffold mis-assembly and the prediction of non-verifiable introns. The ongoing refinement, further analysis and manual curation of the assembled draft genomes, the upcoming information on orthology and paralogy between the *P. sojae* and *P. ramorum* genomes, the closing of still existing gaps in the genomes (Tripathy *et al.*, 2006), and further large-scale EST studies should help to clarify the exact number of gene copies. Unfortunately, it was not possible to infer *P. sojae* and *P. nicotianae* aquaporin gene numbers from the Southern blot analysis using group-specific probes. The probes designed for Group 1 and Group 2 reacted with restriction fragments of the respective other group, however, probes for Group 3, 4 and 5 seemed to be group-specific. As exemplified for *P. nicotianae* aquaporins representing Group 1, restriction fragments labelled in Southern blots may contain several closely clustered genes. Tight clustering of aquaporin genes has been observed in two scaffolds of the *P. sojae* and *P. ramorum* genome, respectively, and has also been shown in the analysis of the *P. nicotianae* genes (i.e. four genes within ~9 kb). Clustering of genes is thought to be a common feature of *Phytophthora* genomes, with genes generally located within high-density gene-islands that are separated by clusters of repetitive sequences (e.g. *PsojNIP* family of *P. sojae*, *ipiO* and *ipiB* of *P. infestans*; in Kamoun, 2003).

Several genes of *P. sojae* and *P. ramorum* displayed ambiguous in-frame start codons. Due to the lack of full-length cDNA information, the most likely translational start site was predicted by promoter analysis. Regions surrounding the start codons in question were analysed for common promoter elements (e.g. TATA-box, CAAT-box, CT-rich regions) and the *Phytophthora*-specific 19 bp motif (Pieterse *et al.*, 1994; McLeod *et al.*,

2004). Furthermore, the G+C content of particular DNA stretches was taken into consideration, as coding-regions of *Phytophthora* genes generally display a higher G+C content than non-coding regions (Win *et al.*, 2006). However, in order to verify these predicted start codons, cDNA information needs to be obtained for the sequences in question by, for example, primer extension experiments or 5' RACE.

P. infestans EST DB analysis revealed a large number of aquaporin gene contigs, however, most of the sequences obtained were very short and were therefore excluded from further analyses. Only two contigs displaying appropriate lengths were retained for multiple sequence alignments and phylogenetic analysis. With the just-released first-draft of the *P. infestans* genome (http://www.broad.mit.edu/annotation/genome/phytophthora_infestans/Home.html), analysis can now be extended and the genome mined for full length aquaporin sequences. However, as the release is only a very rough preliminary assembly with a ~5x sequence coverage, results obtained would have to be viewed with caution.

5.3.2.2 Protein characteristics and conserved domains

Aquaporin proteins within each group and species show high identities and similarities to each other. Inter-species comparison shows good sequence conservation of group members between species when comparing the same group, but generally low identities between members of different groups. The most closely related groups are Group 1 and Group 2 of *P. sojae* and *P. ramorum*, with 52-60% sequence identity. This high level of inter-group identity might explain the cross-reaction in Southern blot analysis of the probes designed to be specific for Group 1 and Group 2. The *P. ramorum* aquaporin PrAq21 shows the highest sequence divergence of all proteins analysed. Clustering with PrAq13 was predicted by both phylogenetic trees (Figure 5.9 and Figure 5.13), however, sequence comparison shows only 40% identity and 59% similarity between the two proteins. Analysis of the multiple sequence alignment of all *Phytophthora* aquaporin sequences (Figure 5.8) by eye indicates that PrAq13 is more similar to Group 2 proteins than to PrAq21. The high sequence divergence of PrAq21 might either be a result of a confined ancient gene duplication or is possibly due to a scaffold mis-assembly resulting in an incorrect protein translation.

A number of highly conserved motifs, located in TMDs and loops B and E, are present throughout all aquaporin sequences. Individual groups displayed different degrees of

divergence within these highly conserved motifs. Group 4 contained the most conserved proteins, whereas Group 5 displayed proteins with the highest number of residue substitutions. Motifs located in TMD 1-4 only display a few residue substitutions, with glycine, threonine and leucine replaced by alanine. The highly conserved glycine in TMD5 is exchanged with either alanine or cysteine in the majority of sequences. This residue is thought to play an import role in the stabilisation of the pore structure by interaction with a second highly conserved glycine in TMD2 (Murata *et al.*, 2000). However, analysis of a large number of aquaporin sequences showed a substitution of the glycine in TMD5 in a number of organisms (Zardoya, 2005), thus is can be inferred that this residue is interchangeable with other residues constituting similar properties without loss of function. In TMD6, the glycine of the motif is commonly replaced by an alanine.

Located in loops B and E, NPA-boxes constitute highly conserved signature motifs of the aquaporin protein family. The proline residues of the two NPA-boxes interact with each other, thus holding the loops together and giving the protein its characteristic shape (Murata *et al.*, 2000). In *Phytophthora* aquaporins, NPA-boxes of loop E are highly conserved. In contrast, in Group 5 four out of nine proteins display a substitution from proline to alanine. Recently, it was shown by sequence analysis that a large number of fungal aquaglyceroporins also display alterations in the loop E NPA box (Pettersson *et al.*, 2005). The signature motif in loop B (SGxHxNPAVT) shows a number of different residue substitutions. In Group 1, the histidine and proline positions are substituted by asparagine and threonine, respectively, in most of the group members. Other groups also display modifications, however, they appear in less relevant positions and less frequently. PrAq21 shows the most divergent motif, with five out of eight positions exchanged. The importance of individual residues of the motif in loop B has been discussed in detail in Chapter 4 (4.3.3).

The so-called selectivity filter, the narrowest constriction of the protein's pore which presumably plays a role in substrate specificity, is formed by four residues (see Chapter 4 Section 4.1.2). Proteins of the aquaporin and GLP cluster display different residues in these positions, thus presenting a simple method to predict the affiliation of proteins which have not been functionally characterised solely by sequence analysis. The selectivity filter residues in *Phytophthora* aquaporins show only a few substitutions.

Generally, three out of four residues correspond to the *E. coli* GlpF residues (W, G, R), suggesting that *Phytophthora* proteins belong to the group of glycerol facilitator proteins and may allow the transport of glycerol and potentially water and other small solutes like urea. However, for the validation of this idea, expression analysis studies in the *Xenopus* oocytes system or functional complementation in yeast need to be performed.

5.3.2.3 Southern blot analysis

As the *P. sojae* and *P. ramorum* genomes contain a large number of aquaporin proteins, it was conceivable that the *P. nicotianae* genome not only contains five genes of Group 1, but also members of other groups. To elucidate this idea, group-specific probes were designed based on multiple sequence alignments of *P. sojae* aquaporin DNA sequences. As *P. ramorum* strains are not available in Australia due to quarantine restrictions, analyses was limited to *P. sojae*. Southern blot analyses, carried out using *P. sojae* and *P. nicotianae* genomic DNA digested with three different restriction enzymes, revealed that the *P. nicotianae* genome indeed also contains genes representing other groups. However, as discussed above, total gene copy numbers could not be determined.

In order to obtain *P. nicotianae* genome information on the remaining aquaporins, the *P. nicotianae* BAC library (Shan and Hardham, 2004) needs to be screened with group-specific probes and further BAC clones isolated and analysed.

5.3.2.4 Phylogenetic analysis

Phylogenetic analysis of *Phytophthora* aquaporins only, and in the context of a large number of aquaporin sequences from other organisms, shows clustering of *Phytophthora* proteins into five groups. All five clades contain orthologous proteins, however, some species-specific gene duplication after speciation resulted in a number of paralogues that do not have orthologous counterparts in the other (*P. sojae* or *P. ramorum*) genome.

The most closely related clade to the *Phytophthora* aquaporins contains aquaporin proteins AQP3, AQP7, AQP9 and AQP10, all of which belong to the GLP cluster and are classified as glycerol transporters. This phylogenetic placement strengthens the idea that *Phytophthora* aquaporins are glycerol facilitator proteins as discussed above.

Chapter 6

General discussion and conclusions

Chapter 6: General discussion and conclusions

Species of the genus *Phytophthora*, class Oomycetes, are able to infect a vast number of important plant species. A few members of the genus are particularly virulent and cause economic and ecological devastation all over the world. *P. nicotianae* and even more so the notorious species *P. infestans*, have an enormous impact on important crop plants, and preventive control measures and losses in production due to pathogen infections amount to billions of dollars per year in the US alone (Erwin and Ribeiro, 1996; Birch and Whisson, 2001). Another species, *P. cinnamomi* which can infect over 2000 plants, not only has a huge impact on economically important crop plants such as avocado, pineapple or cherries, but is also responsible for widespread devastation amongst Australian native plants (Irwin *et al.*, 1995; Erwin and Ribeiro, 1996; Hardham, 2005). *P. cinnamomi* infections have been linked to the destruction of whole ecosystems in Western Australia, Victoria and Tasmania (Cahill, 1993; Irwin *et al.*, 1995; Erwin and Ribeiro, 1996; Hardham, 2005).

With the advances in *Phytophthora*-specific molecular techniques and the emergence of DB comprising vast amounts of EST and genomic data of several *Phytophthora* species in recent years, new avenues have opened for gene discovery and the functional analysis of genes. A number of DB also contain host-plant interaction details, greatly facilitating identification of pathogenicity-specific genes. This aspect is of particular interest, as to date only a few effective chemical fungicides are available. Although destruction by some *Phytophthora* species can largely be contained with these fungicides, a number of resistant isolates have emerged over the years, predominantly in *P. infestans*. Therefore, the identification of potential drug targets that allow for control of these devastating pathogens is a priority for *Phytophthora* research. The search for potential drug targets, as well as the need for detailed annotation and manual curation of genes predicted by computer programs which are deposited in the DB, has led to a sequence-to-phenotype paradigm (Kamoun *et al.*, 2002), which links bioinformatic analysis approaches (i.e. database mining) with functional assays to characterise genes and determine their function and importance in the pathogen. However, reliable transformation techniques required to elucidate gene function still remain a problem in *Phytophthora* research as discussed in 1.3. Thus, alternative methodologies have often been employed for the analysis and characterisation of genes.

The research presented in this study describes the analysis of the single copy gene V-PPase and the aquaporin gene family in the Oomycete *P. nicotianae*. The study encompasses the characterisation of the genes as well as partial functional characterisation of the proteins by means of bioinformatics, immunolocalisation and enzyme assays.

6.1 V-PPases, aquaporins and the WEV

At the outset of this study, the V-PPase and aquaporin genes were considered to possibly play a role in WEV function of *Phytophthora* zoospores. The WEV is an important component of zoospores, which is thought to expel water by a cyclic 'contraction'. Impaired WEV function would lead to swelling of the cell due to continuous uptake of water as the osmolarity inside the cell is much higher than in the surrounding medium, and result in subsequent bursting of the zoospore. This renders the WEV a potential drug target as zoospores are the main infective agents of *Phytophthora* species.

The V-PPase gene was isolated by random screening of a *P. nicotianae* zoospore cDNA library. V-PPases are thought to act as proton pumps to acidify vacuoles and are often co-localised with V-ATPases (e.g. Docampo and Moreno, 2001; Ruiz *et al.*, 2001a; Saliba *et al.*, 2003; Rohloff *et al.*, 2004). This implies that either enzyme can function as a back-up pump for the other or that both enzymes collaboratively facilitate the same function depending on substrate availability. As a V-ATPase was localised to the spongiome of the WEV in *P. nicotianae* zoospores (Mitchell and Hardham, 1999), a co-localisation of the V-PPase protein with the V-ATPase and therefore involvement in WEV function was hypothesised. Results of immunolocalisation studies using antibodies generated in the present study are inconclusive as labelling of the WEV in *P. nicotianae* zoospores may be a result of non-specific antibody binding. However, under the presumption that antibodies indeed bind specifically to the V-PPase protein, immunolocalisation studies suggest that the V-PPase protein localises to a different compartment of the WEV, namely the bladder. This assumed localisation was substantiated by *in vivo* studies with a V-PPase-specific inhibitor which did not impair WEV function. It is likely that the *P. nicotianae* genome contains two V-PPases, as DB analysis of three other *Phytophthora* species revealed the existence of not only one but two V-PPase genes in these genomes, with one gene belonging to the Type I and the other to the Type II group. As antibodies

used for immunlocalisation studies were raised against highly conserved domains that are present in both types of V-PPase, it is likely that the presumed labelling pattern observed represents both V-PPase proteins. Localisation of the V-PPase to the bladder membrane and not the spongiome in the WEV complex would not support a back-up function and/or cooperation in vacuole acidification with the V-ATPase. This and the fact that the V-PPase also occurs in the plasma membrane of the zoospores suggests a different function of the V-PPase protein in *P. nicotianae* zoospores. The V-PPase protein needs to be further characterised in order to fully understand its role. Experiments that may help to elucidate V-PPase function in *Phytophthora* zoospores are outlined in Section 6.3.

The second gene isolated and analysed in the present study belongs to the aquaporin gene family. Aquaporins are water channel proteins which facilitate transport of water through membranes. This characteristic rendered the aquaporins perfect candidates for a role in the WEV complex. The discovery of more than 20 aquaporin proteins in *Phytophthora* genomes was astounding. This situation contrasts with that in the marine diatom *T. pseudonana*, a close relative to *Phytophthora* species and another member of the kingdom Stramenopila, which only has one apparent aquaporin gene in its genome. It is possible, although rather unlikely, that the diatom genome contains MIP-like proteins with highly divergent sequences compared to the orthodox MIPs, which cannot be identified by ordinary BLAST searches. Other protozoan organisms, however, also contain limited numbers of aquaporins. Apicomplexan species, for example, have been shown to possess only one aquaglyceroporin, while species of the order Kinetoplastida display up to five MIPs (Beitz, 2005).

Considering the possible non-specificity of aquaporin-specific antibodies used in immunolocalisation studies, results have to be interpreted cautiously. Under the assumption of specific binding, however, aquaporin antibodies display an identical labelling pattern to the V-PPase antibody, suggesting a co-localisation of aquaporin and V-PPase proteins. Antibodies used for immunolocalisation experiments might be specific to Group 1 aquaporin proteins, as the Y-shaped peptide comprising the consensus C-terminal and N-terminal end of PnAq 1-5 was used for production of polyclonal antibodies in rabbits. N-termini of the other four groups, as well as the C-terminus of Group 2 proteins displayed highly divergent sequences, whereas C-terminal

sequences varied only slightly in Groups 3-5. Depending on antigenicity of the termini in the rabbits, antibodies might have been raised against both termini or a single terminus only. As it is not possible to distinguish the polyclonal antibody population, a cross-reactivity of antibodies between groups can neither be verified nor dismissed. SSCP expression profile analysis showed that four out of five Group 1 genes in *P. nicotianae* are expressed throughout the four asexual developmental stages. No aquaporin transcripts were observed by northern blot analysis in zoospores, however, SSCP data suggest expression of *PnAq2* and *PnAq3*. Thus, proteins of these two genes would constitute aquaporins labelled by antibodies in *P. nicotianae* zoospores. This result suggests a possible involvement of at least some aquaporins in WEV function, however, to clarify the role of individual aquaporin proteins in *Phytophthora* a number of experiments could be carried out as outlined in Section 6.3.

6.2 Synteny of aquaporin genes between *Phytophthora* species

Synteny of genomes describes the preserved order of genes between related species. Chromosomal rearrangements and gene reorganisation are an indicator for the divergence of genomes and may help to evaluate phylogenetic relationships between organisms. Analysis for syntenic regions of genomes is of particular interest when gene families are examined. From phylogenetic analysis and scaffold arrangements of aquaporin genes, synteny between the *P. sojae* and *P. ramorum* genomes was apparent in regions analysed. The tightly clustered aquaporin genes on the large scaffolds 95 and 114 from *P. sojae* and *P. ramorum*, respectively, show identical arrangement of grouped aquaporins and flanking genes on either side. The other two scaffolds containing large numbers of aquaporin genes, 86 from *P. sojae* and 114 from *P. ramorum*, are similar, and appear syntenic, however, some gene re-arrangements have occurred. Due to limited *P. nicotianae* genomic DNA information (i.e. no flanking genes of the aquaporin cluster), syntenic relationships to the other two *Phytophthora* genomes could not be fully resolved. The assembly of super-contigs from existing scaffolds in *P. sojae* and *P. ramorum* is well underway and should give further insight into aquaporin gene arrangements. Additional genomic sequence information should be obtained from *P. nicotianae* in order to compare the genomic regions of interest and determine the degree of synteny between species.

6.3 Future directions

A range of molecular, biochemical and cytological studies will be needed in order to fully understand the role of V-PPases and aquaporin proteins in *P. nicotianae*. As a start, further *P. nicotianae* BAC library screening should be performed. In the case of the V-PPase, a probe could be designed directly from *P. sojae* and *P ramorum* DNA sequences to elucidate the existence of a second, Type I, V-PPase in *P. nicotianae*. To obtain BAC clones that comprise aquaporin proteins of other groups, group-specific probes that were utilised in Southern blot analysis in the present study could be used for screening of the library. In order to obtain detailed expression profiles, real-time PCR experiments with gene-specific primer pairs for the individual aquaporin and V-PPase genes should be considered. To verify the localisation of proteins in zoospores and further address other developmental stages, localisation studies with individual GFP-tagged proteins could be carried out in all four asexual developmental stages of *P. nicotianae*. These experiments would be of particular interest for the aquaporin family, as it might help to understand the existence of the large number of aquaporins in *Phytophthora* species. However, this approach as well as potential gene silencing experiments with dsRNA constructs to evaluate the function of individual genes, will require an efficient transformation protocol for *P. nicotianae*. Although protocols have been vastly improved in our laboratory over the past years, transformation of *P. nicotianae* is still difficult and only a few transformants are obtained in each experiment.

To further elucidate whether V-PPases and/or aquaporins are involved in WEV function in *Phytophthora* zoospores, additional inhibition studies could be carried out. As the substrate AMBP might not inhibit V-PPases efficiently *in vivo* due to its bulky character, it would be worthwhile to consider other options, such as smaller inhibitor substances. Aquaporin-specific inhibitors are also known, for example, silver and gold compounds (Niemietz and Tyerman, 2002) and tetraethylammonium (Brooks *et al.*, 2000). However, inhibition of all functional aquaporins simultaneously might render zoospores or for that matter any given life cycle stage non-viable. Protein specific experiments include a repeat of the pyrophosphatase assay with different buffer systems to evaluate K^+-sensitivity of the V-PPase protein(s) in zoospore microsomal fractions of *P. nicotianae*, and expression studies of individual aquaporin genes in the *Xenopus* oocyte system or yeast complementation assays to elucidate transport capabilities of aquaporin proteins.

In summary, cloning and functional characterisation of the V-PPase and aquaporin genes using a combination of molecular, biochemical, cytological and bioinformatic approaches have revealed a wealth of new information on these gene families in *Phytophthora*. The use of genomic data from a variety of *Phytophthora* species and subsequent bioinformatic analysis has allowed categorisation of individual genes into specific subfamily groups, while biochemical and cytological studies have given some insight into V-PPase and aquaporin gene function. Although further experiments need to be carried out in order to fully characterise these genes, the present study provides a valuable framework for future studies.

Bibliography

Bibliography

Agre, P. (2004). Aquaporin water channels (Nobel Lecture). *Angew Chem Int Ed Engl* **43**(33): 4278-4290

Agre, P. and Kozono, D. (2003). Aquaporin water channels: molecular mechanisms for human diseases. *FEBS Lett* **555**(1): 72-78

Agre, P., Preston, G. M., Smith, B. L., Jung, J. S., Raina, S., Moon, C., Guggino, W. B. and Nielsen, S. (1993a). Aquaporin CHIP: the archetypal molecular water channel. *Am J Physiol* **265**(4): F463-476

Agre, P., Sasaki, S. and Chrispeels, M. J. (1993b). Aquaporins: a family of water channel proteins. *Am J Physiol* **265**(3): F461

Ah Fong, A. M. V. and Judelson, H. S. (2003). Cell cycle regulator Cdc14 is expressed during sporulation but not hyphal growth in the fungus-like oomycete *Phytophthora infestans*. *Mol Microbiol* **50**(2): 487-494

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002). Molecular biology of the cell. New York, Garland Pub.

Alexandersson, E., Fraysse, L., Sjovall-Larsen, S., Gustavsson, S., Fellert, M., Karlsson, M., Johanson, U. and Kjellbom, P. (2005). Whole gene family expression and drought stress regulation of aquaporins. *Plant Mol Biol* **59**(3): 469-484

Allen, R. D. (2000). The contractile vacuole and its membrane dynamics. *Bioessays* **22**(11): 1035-1042

Allen, R. D. and Naitoh, Y. (2002). Osmoregulation and contractile vacuoles of protozoa. *Int Rev Cytol* **215**: 351-394

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17): 3389-3402

Ambikapathy, J., Marshall, J. S., Hocart, C. H. and Hardham, A. R. (2002). The role of proline in osmoregulation in *Phytophthora nicotianae*. *Fungal Genet Biol* **35**(3): 287-299

Anderca, M. I., Suga, S., Furuichi, T., Shimogawara, K., Maeshima, M. and Muto, S. (2004). Functional identification of the glycerol transport activity of *Chlamydomonas reinhardtii* CrMIP1. *Plant Cell Physiol* **45**(9): 1313-1319

Au, K. M., Barabote, R. D., Hu, K. Y. and Saier, M. H., Jr. (2006). Evolutionary appearance of H^+-translocating pyrophosphatases. *Microbiology* **152**(5): 1243-1247

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1993). Current Protocols in Molecular Biology. New York, Greene Pub. Associates.

Baltscheffsky, H., Von Stedingk, L. V., Heldt, H. W. and Klingenberg, M. (1966). Inorganic pyrophosphate: formation in bacterial photophosphorylation. *Science* **153**(740): 1120-1122

Baltscheffsky, M., Nadanaciva, S. and Schultz, A. (1998). A pyrophosphate synthase gene: molecular cloning and sequencing of the cDNA encoding the inorganic pyrophosphate synthase from *Rhodospirillum rubrum*. *Biochim Biophys Acta* **1364**(3): 301-306

Baltscheffsky, M., Schultz, A. and Baltscheffsky, H. (1999). H^+-PPases: a tightly membrane-bound family. *FEBS Lett* **457**(3): 527-533

Barkla, B. J. and Pantoja, O. (1996). Physiology of ion transport across the tonoplast of higher plants. *Annu Rev Plant Physiol Plant Mol Biol* **47**(1): 159-184

Bartnicki-Garcia, S. and Wang, M. C. (1983). Biochemical aspects of morphogenesis in *Phytophthora*, p. 121–137. In: Erwin, D. C., Bartnicki-Garcia, S. and Tsao, P. H. (eds.), *Phytophthora*: Its biology, taxonomy, ecology, and pathology. American Phytopathological Society, St. Paul, Minn.

Baykov, A. A., Bakuleva, N. P. and Rea, P. A. (1993). Steady-state kinetics of substrate hydrolysis by vacuolar H^+-pyrophosphatase. A simple three-state model. *Eur J Biochem* **217**(2): 755-762

Baykov, A. A., Sergina, N. V., Evtushenko, O. A. and Dubnova, E. B. (1996). Kinetic characterization of the hydrolytic activity of the H^+-pyrophosphatase of *Rhodospirillum rubrum* in membrane-bound and isolated states. *Eur J Biochem* **236**(1): 121-127

Becker, B. and Hickisch, A. (2005). Inhibition of contractile vacuole function by brefeldin A. *Plant Cell Physiol* **46**(1): 201-212

Beitz, E. (2005). Aquaporins from pathogenic protozoan parasites: structure, function and potential for chemotherapy. *Biol Cell* **97**(6): 373-383

Belogurov, G. A. and Lahti, R. (2002). A lysine substitute for K^+. A460K mutation eliminates K^+ dependence in H^+-pyrophosphatase of *Carboxydothermus hydrogenoformans*. *J Biol Chem* **277**(51): 49651-49654

Belogurov, G. A., Turkina, M. V., Penttinen, A., Huopalahti, S., Baykov, A. A. and Lahti, R. (2002). H^+-pyrophosphatase of *Rhodospirillum rubrum*. High yield expression in *Escherichia coli* and identification of the Cys residues responsible for inactivation my mersalyl. *J Biol Chem* **277**(25): 22209-22214

Benga, G. (2003). Birth of water channel proteins-the aquaporins. *Cell Biol Int* **27**(9): 701-709

Benga, G., Popescu, O., Borza, V., Pop, V. I., Muresan, A., Mocsy, I., Brain, A. and Wrigglesworth, J. M. (1986a). Water permeability in human erythrocytes: identification of membrane proteins involved in water transport. *Eur J Cell Biol* **41**(2): 252-262

Benga, G., Popescu, O., Pop, V. I. and Holmes, R. P. (1986b). p-(Chloromercuri)benzenesulfonate binding by membrane proteins and the inhibition of water transport in human erythrocytes. *Biochemistry* **25**(7): 1535-1538

Berry, V., Francis, P., Kaushal, S., Moore, A. and Bhattacharya, S. (2000). Missense mutations in MIP underlie autosomal dominant 'polymorphic' and lamellar cataracts linked to 12q. *Nature Genet* **25**(1): 15-17

Bhattacharjee, S., Hiller, N. L., Liolios, K., Win, J., Kanneganti, T. D., Young, C., Kamoun, S. and Haldar, K. (2006). The malarial host-targeting signal is conserved in the Irish potato famine pathogen. *PLoS Pathog* **2**(5): e50

Bio-Rad handbook, Garfin, D., Heerdt, L. (eds.) 2-D electrophoresis for proteomics - A methods and product manual.

Birch, P. R. J. and Whisson, S. C. (2001). *Phytophthora infestans* enters the genomics era. *Mol Plant Pathol* **2**(5): 257-263

Blanco, F. A. and Judelson, H. S. (2005). A bZIP transcription factor from *Phytophthora* interacts with a protein kinase and is required for zoospore motility and plant infection. *Mol Microbiol* **56**(3): 638-648

Bonhivers, M., Carbrey, J. M., Gould, S. J. and Agre, P. (1998). Aquaporins in *Saccharomyces*. Genetic and functional distinctions between laboratory and wild-type strains. *J Biol Chem* **273**(42): 27565-27572

Borgnia, M. J. and Agre, P. (2001). Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from *Escherichia coli*. *Proc Natl Acad Sci U S A* **98**(5): 2888-2893

Borgnia, M. J., Kozono, D., Calamita, G., Maloney, P. C. and Agre, P. (1999). Functional reconstitution and characterization of AqpZ, the *E. coli* water channel protein. *J Mol Biol* **291**(5): 1169-1179

Bottin, A., Larche, L., Villalba, F., Gaulin, E., Esquerre-Tugaye, M. T. and Rickauer, M. (1999). Green fluorescent protein (GFP) as gene expression reporter and vital marker for studying development and microbe-plant interaction in the tobacco pathogen *Phythphthora parasitica* var. *nicotianae*. *FEMS Microbiol Lett* **176**(1): 51-56

Bouck, G. B. (1972). Architecture and assembly of mastigonemes, p. 237-276. In: Du Praw, E. J. (eds.), Advances in cell and molecular biology, Vol. 2. Academic Press, New York

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248-254

Braun, T., Philippsen, A., Wirtz, S., Borgnia, M. J., Agre, P., Kuhlbrandt, W., Engel, A. and Stahlberg, H. (2000). The 3.7 A projection map of the glycerol facilitator GlpF: a variant of the aquaporin tetramer. *EMBO Rep* **1**(2): 183-189

Brightman, A. O., Navas, P., Minnifield, N. M. and Morre, D. J. (1992). Pyrophosphate-induced acidification of trans cisternal elements of rat liver Golgi apparatus. *Biochim Biophys Acta* **1104**(1): 188-194

Brini, F., Gaxiola, R. A., Berkowitz, G. A. and Masmoudi, K. (2005). Cloning and characterization of a wheat vacuolar cation/proton antiporter and pyrophosphatase proton pump. *Plant Physiol Biochem* **43**(4): 347-354

Britten, C. J., Turner, J. C. and Rea, P. A. (1989). Identification and purification of substrate-binding subunit of higher plant H^+-translocating inorganic pyrophosphatase. *FEBS Letters* **256**(1-2): 200

Brooks, H. L., Regan, J. W. and Yool, A. J. (2000). Inhibition of aquaporin-1 water permeability by tetraethylammonium: involvement of the loop E pore region. *Mol Pharmacol* **57**(5): 1021-1026

Bryant, D., Cummins, I., Dixon, D. P. and Edwards, R. (2006). Cloning and characterization of a theta class glutathione transferase from the potato pathogen *Phytophthora infestans*. *Phytochemistry* **67**(14): 1427-1434

Caetano-Anollés, G. and Gresshoff, P. (1994). Staining Nucleic Acids with Silver: An Alternative to Radioisotopic and Fluorescent Labeling. *Promega Notes Magazine* **45**: 13-20

Caffrey, M. (2003). Membrane protein crystallization. *J Struct Biol* **142**(1): 108-132

Cahill, D. M. (1993). Review of *Phytophthora* diseases in Australia. Canberra, Rural Industries Research and Development Corporation, Report Series No. 93/94.

Cahill, D. M., Cope, M. and Hardham, A. R. (1996). Thrust reversal by tubular mastigonemes: immunological evidence for a role of mastigonemes in forward motion of zoospores of *Phytophthora cinnamomi*. *Protoplasma* **194**: 18-28

Cahill, D. M. and Hardham, A. R. (1994). Exploitation of zoospore taxis in the development of a novel dipstick immunoassay for the specific detection of *Phytophthora cinnamomi*. *Phytopathology* **84**(2): 193-200

Calamita, G. (2000). The *Escherichia coli* aquaporin-Z water channel. *Mol Microbiol* **37**(2): 254-262

Campanella, J. J., Bitincka, L. and Smalley, J. (2003). MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. *BMC Bioinformatics* **4**: 29-32

Carbrey, J. M., Bonhivers, M., Boeke, J. D. and Agre, P. (2001). Aquaporins in *Saccharomyces*: Characterization of a second functional water channel protein. *Proc Natl Acad Sci U S A* **98**(3): 1000-1005

Carlile, M. J. (1983). Motility, taxis and tropoism in *Phytophthora*, p. 95-107. In: Erwin, D. C., Bartnicki-Garcia, S. and Tsao, P. H. (eds.), *Phytophthora*: Its biology, taxonomy, ecology, and pathology. American Phytopathological Society, St. Paul, Minn

Cavalier-Smith, T. (1993). Kingdom protozoa and its 18 phyla. *Microbiol Rev* **57**(4): 953-994

Chanson, A., Fichmann, J., Spear, D. and Taiz, L. (1985). Pyrophosphate-Driven Proton Transport by Microsomal Membranes of Corn Coleoptiles. *Plant Physiol* **79**(1): 159-164

Chaumont, F., Barrieu, F., Wojcik, E., Chrispeels, M. J. and Jung, R. (2001). Aquaporins constitute a large and highly divergent protein family in maize. *Plant Physiol* **125**(3): 1206-1215

Cheng, A., van Hoek, A. N., Yeager, M., Verkman, A. S. and Mitra, A. K. (1997). Three-dimensional organization of a human water channel. *Nature* **387**(6633): 627-630

Cho, C. W. and Fuller, M. S. (1989). Observations of the water expulsion vacuole of *Phytophthora palmivora*. *Protoplasma* **149**: 47-55

Claros, M. G. and von Heijne, G. (1994). TopPred II: an improved software for membrane protein structure predictions. *Comput Appl Biosci* **10**(6): 685-686

Cooke, D. E., Drenth, A., Duncan, J. M., Wagels, G. and Brasier, C. M. (2000). A molecular phylogeny of *Phytophthora* and related oomycetes. *Fungal Genet Biol* **30**(1): 17-32

Cooperman, B. S. (1982). The mechanism of action of yeast inorganic pyrophosphatase. *Methods Enzymol* **87**: 526-548

Cooperman, B. S., Baykov, A. A. and Lahti, R. (1992). Evolutionary conservation of the active site of soluble inorganic pyrophosphatase. *Trends Biochem Sci* **17**(7): 262-266

Costanzo, S., Ospina-Giraldo, M. D., Deahl, K. L., Baker, C. J. and Jones, R. W. (2006). Gene duplication event in family 12 glycosyl hydrolase from *Phytophthora* spp. *Fungal Genet Biol* **Epub** doi:10.1016/j.fgb.2006.04.006

Cserzo, M., Wallin, E., Simon, I., von Heijne, G. and Elofsson, A. (1997). Prediction of transmembrane α-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng* **10**(6): 673-676

Cvitanich, C. and Judelson, H. S. (2003a). A gene expressed during sexual and asexual sporulation in *Phytophthora infestans* is a member of the Puf family of translational regulators. *Eukaryot Cell* **2**(3): 465-473

Cvitanich, C. and Judelson, H. S. (2003b). Stable transformation of the oomycete, *Phytophthora infestans*, using microprojectile bombardment. *Curr Genet* **42**(4): 228-235

Daniels, M. J., Chrispeels, M. J. and Yeager, M. (1999). Projection structure of a plant vacuole membrane aquaporin by electron cryo-crystallography. *J Mol Biol* **294**(5): 1337-1349

de Groot, B. L. and Grubmuller, H. (2001). Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science* **294**(5550): 2353-2357

Deen, P. M., Verdijk, M. A., Knoers, N. V., Wieringa, B., Monnens, L. A., van Os, C. H. and van Oost, B. A. (1994). Requirement of human renal water channel aquaporin-2 for vasopressin-dependent concentration of urine. *Science* **264**(5155): 92-95

Denker, B. M., Smith, B. L., Kuhajda, F. P. and Agre, P. (1988). Identification, purification, and partial characterization of a novel Mr 28,000 integral membrane protein from erythrocytes and renal tubules. *J Biol Chem* **263**(30): 15634-15642

Docampo, R. and Moreno, S. N. (2001). The acidocalcisome. *Mol Biochem Parasitol* **114**(2): 151-159

Dong, W. B., Latijnhouwers, M., Jiang, R. H. Y., Meijer, H. J. G. and Govers, F. (2004). Downstream targets of the *Phytophthora infestans* Gα subunit PiGPA1 revealed by cDNA-AFLP. *Mol Plant Pathol* **5**(5): 483-494

Drozdowicz, Y. M., Kissinger, J. C. and Rea, P. A. (2000). AVP2, a sequence-divergent, K^+-insensitive H^+-translocating inorganic pyrophosphatase from *Arabidopsis*. *Plant Physiol* **123**(1): 353-362

Drozdowicz, Y. M., Lu, Y. P., Patel, V., Fitz-Gibbon, S., Miller, J. H. and Rea, P. A. (1999). A thermostable vacuolar-type membrane pyrophosphatase from the archaeon *Pyrobaculum aerophilum*: implications for the origins of pyrophosphate-energized pumps. *FEBS Lett* **460**(3): 505-512

Drozdowicz, Y. M. and Rea, P. A. (2001). Vacuolar H^+ pyrophosphatases: from the evolutionary backwaters into the mainstream. *Trends Plant Sci* **6**(5): 206-211

Drozdowicz, Y. M., Shaw, M., Nishi, M., Striepen, B., Liwinski, H. A., Roos, D. S. and Rea, P. A. (2003). Isolation and characterization of TgVP1, a type I vacuolar H^+-translocating pyrophosphatase from *Toxoplasma gondii*. The dynamics of its subcellular localization and the cellular effects of a diphosphonate inhibitor. *J Biol Chem* **278**(2): 1075-1085

Duncan, J. M. (1999). *Phytophthora*- an abiding threat to our crops. *Microbiology Today* **26**: 114 - 116

Duniway, J. M. (1976). Movement of zoospores of *Phytophthora cryptogea* in soils of various textures and matri potentials. *Phytopathology* **66**(7): 877-882

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics* **14**(9): 755-763

El Karkouri, K., Gueune, H. and Delamarche, C. (2005). MIPDB: a relational database dedicated to MIP family proteins. *Biol Cell* **97**(7): 535-543

Erwin, D. C. and Ribeiro, O. K. (1996). *Phytophthora* diseases worldwide. St. Paul, Minn., APS Press.

Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**: 164-166

Finbow, M. E. and Harrison, M. A. (1997). The vacuolar H^+-ATPase: a universal proton pump of eukaryotes. *Biochem J* **324**: 697-712

Flick, K. M., Shaulsky, G. and Loomis, W. F. (1997). The *wacA* gene of *Dictyostelium discoideum* is a developmentally regulated member of the MIP family. *Gene* **195**(2): 127-130

Fok, A. K., Aihara, M. S., Ishida, M., Nolta, K. V., Steck, T. L. and Allen, R. D. (1995). The pegs on the decorated tubules of the contractile vacuole complex of *Paramecium* are proton pumps. *J Cell Sci* **108**: 3163-3170

Fortin, M. G., Morrison, N. A. and Verma, D. P. (1987). Nodulin-26, a peribacteroid membrane nodulin is expressed independently of the development of the peribacteroid compartment. *Nucleic Acids Res* **15**(2): 813-824

Francis, P., Chung, J. J., Yasui, M., Berry, V., Moore, A., Wyatt, M. K., Wistow, G., Bhattacharya, S. S. and Agre, P. (2000). Functional impairment of lens aquaporin in two families with dominantly inherited cataracts. *Hum Mol Genet* **9**(15): 2329-2334

Froger, A., Rolland, J. P., Bron, P., Lagree, V., Le Caherec, F., Deschamps, S., Hubert, J. F., Pellerin, I., Thomas, D. and Delamarche, C. (2001). Functional characterization of a microbial aquaglyceroporin. *Microbiology* **147**: 1129-1135

Froger, A., Tallur, B., Thomas, D. and Delamarche, C. (1998). Prediction of functional residues in water channels and related proteins. *Protein Sci* **7**(6): 1458-1468

Fu, D., Libson, A., Miercke, L. J. W., Weitzman, C., Nollert, P., Krucinski, J. and Stroud, R. M. (2000). Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290**(5491): 481-486

Fujiyoshi, Y., Mitsuoka, K., de Groot, B. L., Philippsen, A., Grubmuller, H., Agre, P. and Engel, A. (2002). Structure and function of water channels. *Curr Opin Struct Biol* **12**(4): 509-515

Fukuda, A., Chiba, K., Maeda, M., Nakamura, A., Maeshima, M. and Tanaka, Y. (2004). Effect of salt and osmotic stresses on the expression of genes for the vacuolar H^+-pyrophosphatase, H^+-ATPase subunit A, and Na^+/H^+ antiporter from barley. *J Exp Bot* **55**(397): 585-594

Gajendran, K., Gonzales, M. D., Farmer, A., Archuleta, E., Win, J., Waugh, M. E. and Kamoun, S. (2006). *Phytophthora* functional genomics database (PFGD): functional genomics of phytophthora-plant interactions. *Nucleic Acids Res* **34**: D465-470

Gasser, R. B. and Monti, J. R. (1997). Identification of parasitic nematodes by PCR-SSCP of ITS-2 rDNA. *Mol Cell Probes* **11**(3): 201-209

Gaulin, E., Jauneau, A., Villalba, F., Rickauer, M., Esquerre-Tugaye, M. T. and Bottin, A. (2002). The CBEL glycoprotein of *Phytophthora parasitica* var. *nicotianae* is involved in cell wall deposition and adhesion to cellulosic substrates. *J Cell Sci* **115**: 4565-4575

Ghosh, K., Cappiello, C. D., McBride, S. M., Occi, J. L., Cali, A., Takvorian, P. M., McDonald, T. V. and Weiss, L. M. (2005). Functional characterization of a putative aquaporin from *Encephalitozoon cuniculi*, a microsporidia pathogenic to humans. *Int J Parasitol* **36**(1): 57-62

Gonen, T., Sliz, P., Kistler, J., Cheng, Y. and Walz, T. (2004). Aquaporin-0 membrane junctions reveal the structure of a closed water pore. *Nature* **429**(6988): 193-197

Gordon-Weeks, R., Parmar, S., Davies, T. G. and Leigh, R. A. (1999). Structural aspects of the effectiveness of bisphosphonates as competitive inhibitors of the plant vacuolar proton-pumping pyrophosphatase. *Biochem J* **337**: 373-377

Gordon-Weeks, R., Steele, S. H. and Leigh, R. A. (1996). The role of magnesium, pyrophosphate, and their complexes as substrates and activators of the vacuolar H^+-pumping inorganic pyrophosphatase. *Plant Physiol* **111**(1): 195-202

Götesson, A., Marshall, J., Jones, D. and Hardham, A. (2002). Characterization and evolutionary analysis of a large polygalacturonase gene family in the oomycete plant pathogen *Phytophthora cinnamomi*. *Mol Plant-Microbe Interact* **15**(9): 907-921

Grenville-Briggs, L. J., Avrova, A. O., Bruce, C. R., Williams, A., Whisson, S. C., Birch, P. R. and van West, P. (2005). Elevated amino acid biosynthesis in *Phytophthora infestans* during appressorium formation and potato infection. *Fungal Genet Biol* **42**(3): 244-256

Guan, L., Smirnova, I. N., Verner, G., Nagamori, S. and Kaback, H. R. (2006). Manipulating phospholipids for crystallization of a membrane transport protein. *Proc Natl Acad Sci U S A* **103**(6): 1723-1726

Guarro, J., GeneJ and Stchigel, A. M. (1999). Developments in fungal taxonomy. *Clin Microbiol Rev* **12**(3): 454-500

Gubler, F. and Hardham, A. R. (1988). Secretion of adhesive material during encystment of *Phytophthora cinnamomi* zoospores, characterized by immunogold labeling with monoclonal-antibodies to components of peripheral vesicles. *J Cell Sci* **90**: 225-235

Gubler, F. and Hardham, A. R. (1990). Protein storage in large peripheral vesicles in *Phytophthora* zoospores and its breakdown after cyst germination. *Exp Mycol* **14**(4): 393-404

Gubler, F., Hardham, A. R. and Duniec, J. (1989). Characterizing adhesiveness of *Phytophthora cinnamomi* zoospores during encystment. *Protoplasma* **149**(1): 24-30

Guo, S., Yin, H., Zhang, X., Zhao, F., Li, P., Chen, S., Zhao, Y. and Zhang, H. (2006). Molecular cloning and characterization of a vacuolar H^+-pyrophosphatase gene, *SsVP*, from the halophyte *Suaeda salsa* and its overexpression increases salt and drought tolerance of *Arabidopsis*. *Plant Mol Biol* **60**(1): 41-50

Hall, G. (1993). An integrated approach to the analysis of variation in *Phytophthora nicotianae* and a redescription of the species. *Mycol Res* **97**: 559-574

Hall, T. A. (1999). BioEdit:a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**: 95-98

Hansen, M., Kun, J. F., Schultz, J. E. and Beitz, E. (2002). A single, bi-functional aquaglyceroporin in blood-stage *Plasmodium falciparum* malaria parasites. *J Biol Chem* **277**(7): 4874-4882

Hardham, A. R. (1985). Studies on the cell surface of zoospores and cysts of the fungus *Phytophthora cinnamomi*: The influence of fixation on patterns of lectin binding. *J Histochem Cytochem* **33**(2): 110-118

Hardham, A. R. (1987). Ultrastructure and serial section reconstruction of zoospores of the fungus *Phytophthora cinnamomi*. *Exp Mycol* **11**(4): 297-306

Hardham, A. R. (2005). *Phytophthora cinnamomi*. *Mol Plant Pathol* **6**(6): 589-604

Hardham, A. R., Cahill, D. M., Cope, M., Gabor, B. K., Gubler, F. and Hyde, G. J. (1994). Cell-surface antigens of *Phytophthora* spores - Biological and taxonomic characterization. *Protoplasma* **181**(1-4): 213-232

Hardham, A. R. and Gubler, F. (1990). Polarity of attachment of zoospores of a root pathogen and pre-alignment of the emerging germ tube. *Cell Biol Int* **14**(11): 947-956

Hardham, A. R., Gubler, F., Duniec, J. and Elliott, J. (1991). A review of methods for the production and use of monoclonal antibodies to study zoosporic plant pathogens. *J Microsc* **162**: 305-318

Hardham, A. R. and Hyde, G. J. (1997). Asexual sporulation in the Oomycetes. *Adv Bot Res* **24**: 353-398

Harries, W. E., Akhavan, D., Miercke, L. J., Khademi, S. and Stroud, R. M. (2004). The channel architecture of aquaporin 0 at a 2.2-A resolution. *Proc Natl Acad Sci U S A* **101**(39): 14045-14050

Hemmes, D. E. (1983). Cytology of *Phytophthora*, p. 9-40. In: Erwin, D. C., Bartnicki-Garcia, S. and Tsao, P. H. (eds.), *Phytophthora*: Its biology, taxonomy, ecology, and pathology. American Phytopathological Society, St. Paul, Minn

Heuser, J., Zhu, Q. and Clarke, M. (1993). Proton pumps populate the contractile vacuoles of *Dictyostelium* amoebae. *J Cell Biol* **121**(6): 1311-1327

Hill, A. E., Shachar-Hill, B. and Shachar-Hill, Y. (2004). What are aquaporins for? *J Membr Biol* **197**(1): 1-32

Hill, J. E., Scott, D. A., Luo, S. and Docampo, R. (2000). Cloning and functional expression of a gene encoding a vacuolar-type proton-translocating pyrophosphatase from *Trypanosoma cruzi*. *Biochem J* **351**: 281-288

Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G. and Blanquet, S. (1989). Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc Natl Acad Sci U S A* **86**(21): 8247-8251

Hirokawa, T., Boon-Chieng, S. and Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**(4): 378-379

Hirono, M., Mimura, H., Nakanishi, Y. and Maeshima, M. (2005). Expression of functional *Streptomyces coelicolor* H^+-pyrophosphatase and characterization of its molecular properties. *J Biochem (Tokyo)* **138**(2): 183-191

Hofmann, K. and Stoffel, W. (1993). TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374**: 166

Hohmann, I., Bill, R. M., Kayingo, I. and Prior, B. A. (2000). Microbial MIP channels. *Trends Microbiol* **8**(1): 33-38

Holder, M. and Lewis, P. O. (2003). Phylogeny estimation: traditional and Bayesian approaches. *Nature Rev Genet* **4**(4): 275-284

Hsiao, Y. Y., Van, R. C., Hung, H. H. and Pan, R. L. (2002). Diethylpyrocarbonate inhibition of vacuolar H^+-pyrophosphatase possibly involves a histidine residue. *J Protein Chem* **21**(1): 51-58

Hsiao, Y. Y., Van, R. C., Hung, S. H., Lin, H. H. and Pan, R. L. (2004). Roles of histidine residues in plant vacuolar H^+-pyrophosphatase. *Biochim Biophys Acta* **1608**(2-3): 190-199

Ikeda, M., Beitz, E., Kozono, D., Guggino, W. B., Agre, P. and Yasui, M. (2002a). Characterization of aquaporin-6 as a nitrate channel in mammalian cells. Requirement of pore-lining residue threonine 63. *J Biol Chem* **277**(42): 39873-39879

Ikeda, M., Rahman, M., Moritani, C., Umami, K., Tanimura, Y., Akagi, R., Tanaka, Y., Maeshima, M. and Watanabe, Y. (1999). Gene note. A vacuolar H^+-pyrophosphatase in *Acetabularia acetabulum*: molecular cloning and comparison with higher plants and a bacterium. *J Exp Bot* **50**(330): 139-140

Ikeda, M., Umami, K., Hinohara, M., Tanimura, Y., Ohmae, A., Nakanishi, Y. and Maeshima, M. (2002b). Functional expression of *Acetabularia acetabulum* vacuolar H^+-pyrophosphatase in a yeast VMA3-deficient strain. *J Exp Bot* **53**(378): 2273-2275

Industry Commission (1994). The tobacco growing and manufacturing industries. Canberra, Australian Government Publishing Service.

Inman, J. T., Flores, H. R., May, G. D., Weller, J. W. and Bell, C. J. (2001). A high-throughput distributed DNA sequence analysis and database system. *Ibm Systems Journal* **40**(2): 464-486

Irwin, J. A. G., Cahill, D. M. and Drenth, A. (1995). *Phytophthora* in Australia. *Australian Journal of Agricultural Research* **46**(7): 1311-1337

Ishibashi, K. (2006). Aquaporin subfamily with unusual NPA boxes. *Biochim Biophys Acta* **Epub** doi:10.1016/j.bbamem.2006.02.024

Ishibashi, K., Kuwahara, M., Gu, Y., Kageyama, Y., Tohsaka, A., Suzuki, F., Marumo, F. and Sasaki, S. (1997). Cloning and functional expression of a new water channel abundantly expressed in the testis permeable to water, glycerol, and urea. *J Biol Chem* **272**(33): 20782-20786

Ishibashi, K., Kuwahara, M., Kageyama, Y., Sasaki, S., Suzuki, M. and Imai, M. (2000). Molecular cloning of a new aquaporin superfamily in mammals: AQPX1 and AQPX2, p. 123-126. In: Hohmann, S. and Nielsen, S. (eds.), Molecular biology and physiology of water and solute transport. Kluwer Academic/Plenum Publishers, New York

Ishikawa, F., Suga, S., Uemura, T., Sato, M. H. and Maeshima, M. (2005). Novel type aquaporin SIPs are mainly localized to the ER membrane and show cell-specific expression in *Arabidopsis thaliana*. *FEBS Lett* **579**(25): 5814-5820

Itoh, T., Rai, T., Kuwahara, M., Ko, S. B., Uchida, S., Sasaki, S. and Ishibashi, K. (2005). Identification of a novel aquaporin, AQP12, expressed in pancreatic acinar cells. *Biochem Biophys Res Commun* **330**(3): 832-838

Jensen, M. O., Tajkhorshid, E. and Schulten, K. (2001). The mechanism of glycerol conduction in aquaglyceroporins. *Structure* **9**(11): 1083-1093

Jiang, J., Daniels, B. V. and Fu, D. (2006a). Crystal structure of AqpZ tetramer reveals two distinct Arg-189 conformations associated with water permeation through the narrowest constriction of the water-conducting channel. *J Biol Chem* **281**(1): 454-460

Jiang, R. H., Dawe, A. L., Weide, R., van Staveren, M., Peters, S., Nuss, D. L. and Govers, F. (2005). Elicitin genes in *Phytophthora infestans* are clustered and interspersed with various transposon-like elements. *Mol Genet Genomics* **273**(1): 20-32

Jiang, R. H., Tyler, B. M., Whisson, S. C., Hardham, A. R. and Govers, F. (2006b). Ancient origin of elicitin gene clusters in *Phytophthora* genomes. *Mol Biol Evol* **23**(2): 338-351

Jiang, S. S., Fan, L. L., Yang, S. J., Kuo, S. Y. and Pan, R. L. (1997). Purification and characterization of thylakoid membrane-bound inorganic pyrophosphatase from *Spinacia oleracia* L. *Arch Biochem Biophys* **346**(1): 105-112

Johanson, U. and Gustavsson, S. (2002). A new subfamily of major intrinsic proteins in plants. *Mol Biol Evol* **19**(4): 456-461

Johanson, U., Karlsson, M., Johansson, I., Gustavsson, S., Sjovall, S., Fraysse, L., Weig, A. R. and Kjellbom, P. (2001). The complete set of genes encoding major intrinsic proteins in *Arabidopsis* provides a framework for a new nomenclature for major intrinsic proteins in plants. *Plant Physiol* **126**(4): 1358-1369

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**(2): 195-202

Judelson, H. S. (1993). Intermolecular ligation mediates efficient cotransformation in *Phytophthora infestans*. *Mol Gen Genet* **239**(1-2): 241-250

Judelson, H. S. (1997). Expression and inheritance of sexual preference and selfing potential in *Phytophthora infestans*. *Fungal Genet Biol* **21**(2): 188-197

Judelson, H. S. and Blanco, F. A. (2005). The spores of *Phytophthora*: weapons of the plant destroyer. *Nature Rev Microbiol* **3**(1): 47-58

Judelson, H. S., Coffey, M. D., Arredondo, F. R. and Tyler, B. M. (1993a). Transformation of the oomycete pathogen *Phytophthora megasperma* f. sp. *glycinea* occurs by DNA integration into single or multiple chromosomes. *Curr Genet* **23**(3): 211-218

Judelson, H. S., Dudler, R., Pieterse, C. M., Unkles, S. E. and Michelmore, R. W. (1993b). Expression and antisense inhibition of transgenes in *Phytophthora infestans* is modulated by choice of promoter and position effects. *Gene* **133**(1): 63-69

Judelson, H. S. and Randall, T. A. (1998). Families of repeated DNA in the oomycete *Phytophthora infestans* and their distribution within the genus. *Genome* **41**(4): 605-615

Judelson, H. S., Tyler, B. M. and Michelmore, R. W. (1991). Transformation of the oomycete pathogen, *Phytophthora infestans*. *Mol Plant-Microbe Interact* **4**(6): 602-607

Judelson, H. S. and Whittaker, S. L. (1995). Inactivation of transgenes in *Phytophthora infestans* is not associated with their deletion, methylation, or mutation. *Curr Genet* **28**(6): 571-579

Jung, J. S., Preston, G. M., Smith, B. L., Guggino, W. B. and Agre, P. (1994). Molecular structure of the water channel through aquaporin CHIP. The hourglass model. *J Biol Chem* **269**(20): 14648-14654

Juretic, D., Zoranic, L. and Zucic, D. (2002). Basic charge clusters and predictions of membrane protein topology. *J Chem Inf Comput Sci* **42**(3): 620-632

Kaldenhoff, R. and Fischer, M. (2006a). Aquaporins in plants. *Acta Physiol (Oxf)* **187**(1-2): 169-176

Kaldenhoff, R. and Fischer, M. (2006b). Functional aquaporin diversity in plants. *Biochim Biophys Acta* **Epub** doi:10.1016/j.bbamem.2006.03.012

Kamoun, S. (2003). Molecular genetics of pathogenic oomycetes. *Eukaryot Cell* **2**(2): 191-199

Kamoun, S., Dong, S., Hamada, W., Huitema, E., Kinney, D., Morgan, W. R., Styer, A., Testa, A. and Torto, T. A. (2002). From sequence to phenotype: Functional genomics of *Phytophthora*. *Can J Plant Path* **24**: 6-9

Kamoun, S., Hraber, P., Sobral, B., Nuss, D. and Govers, F. (1999). Initial assessment of gene diversity for the oomycete pathogen *Phytophthora infestans* based on expressed sequences. *Fungal Genet Biol* **28**(2): 94-106

Kamoun, S., van West, P. and Govers, F. (1998a). Quantification of late blight resistance of potato using transgenic *Phytophthora infestans* expressing beta-glucuronidase. *Eur J Plant Pathol* **104**(5): 521-525

Kamoun, S., van West, P., Vleeshouwers, V. G., de Groot, K. E. and Govers, F. (1998b). Resistance of *Nicotiana benthamiana* to *Phytophthora infestans* is mediated by the recognition of the elicitor protein INF1. *Plant Cell* **10**(9): 1413-1426

Karpel, J. E. and Bisson, L. F. (2006). Aquaporins in *Saccharomyces cerevisiae* wine yeast. *FEMS Microbiol Lett* **257**(1): 117-123

Kayingo, G., Bill, R. M., Calamita, G., Hohmann, S. and Prior, B. A. (2001). Microbial water channels and glycerol facilitators, p. 335-370. In: Benos, D., Simon, S., Hohmann, S., Agre, P. and Nielsen, S. (eds.), Aquaporins - Current Topics in Membranes, Vol. 51. Academic Press, San Diego

Khew, K. L. and Zentmyer, G. A. (1973). Chemotactic response of zoospores of 5 species of *Phytophthora*. *Phytopathology* **63**(12): 1511-1517

Kim, E. J., Zhen, R. G. and Rea, P. A. (1994a). Heterologous expression of plant vacuolar pyrophosphatase in yeast demonstrates sufficiency of the substrate-binding subunit for proton transport. *Proc Natl Acad Sci U S A* **91**(13): 6128-6132

Kim, E. J., Zhen, R. G. and Rea, P. A. (1995). Site-directed mutagenesis of vacuolar H^+-pyrophosphatase. Necessity of Cys634 for inhibition by maleimides but not catalysis. *J Biol Chem* **270**(6): 2630-2635

Kim, Y., Kim, E. J. and Rea, P. A. (1994b). Isolation and characterization of cDNAs encoding the vacuolar H^+-pyrophosphatase of *Beta vulgaris*. *Plant Physiol* **106**(1): 375-382

King, L. S., Choi, M., Fernandez, P. C., Cartron, J. P. and Agre, P. (2001). Defective urinary-concentrating ability due to a complete deficiency of aquaporin-1. *N Engl J Med* **345**(3): 175-179

King, L. S., Kozono, D. and Agre, P. (2004). From structure to disease: the evolving tale of aquaporin biology. *Nature Rev Mol Cell Biol* **5**(9): 687-698

King, L. S. and Yasui, M. (2002). Aquaporins and disease: lessons from mice to humans. *Trends Endocrinol Metab* **13**(8): 355-360

Kishida, K., Shimomura, I., Kondo, H., Kuriyama, H., Makino, Y., Nishizawa, H., Maeda, N., Matsuda, M., Ouchi, N., Kihara, S., Kurachi, Y., Funahashi, T. and Matsuzawa, Y. (2001). Genomic structure and insulin-mediated repression of the aquaporin adipose (AQPap), adipose-specific glycerol channel. *J Biol Chem* **276**(39): 36251-36260

Knight, H., Trewavas, A. J. and Knight, M. R. (1996). Cold calcium signaling in *Arabidopsis* involves two cellular pools and a change in calcium signature after acclimation. *Plant Cell* **8**(3): 489-503

Knight, J. (2002). Fears mount as oak blight infects redwoods. *Nature* **415**(6869): 251

Kobae, Y., Mizutani, M., Segami, S. and Maeshima, M. (2006). Immunochemical analysis of aquaporin isoforms in *Arabidopsis* suspension-cultured cells. *Biosci Biotechnol Biochem* **70**(4): 980-987

Kozono, D., Ding, X., Iwasaki, I., Meng, X., Kamagata, Y., Agre, P. and Kitagawa, Y. (2003). Functional expression and characterization of an archaeal aquaporin. AqpM from *Methanothermobacter marburgensis*. *J Biol Chem* **278**(12): 10649-10656

Kozono, D., Yasui, M., King, L. S. and Agre, P. (2002). Aquaporin water channels: atomic structure molecular dynamics meet clinical medicine. *J Clin Invest* **109**(11): 1395-1399

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**(3): 567-580

Kroon, L. P., Bakker, F. T., van den Bosch, G. B., Bonants, P. J. and Flier, W. G. (2004). Phylogenetic analysis of *Phytophthora* species based on mitochondrial and nuclear DNA sequences. *Fungal Genet Biol* **41**(8): 766-782

Kruse, E., Uehlein, N. and Kaldenhoff, R. (2006). The aquaporins. *Genome Biol* 7(2): 206

Kukita, Y., Tahira, T., Sommer, S. S. and Hayashi, K. (1997). SSCP analysis of long DNA fragments in low pH gel. *Hum Mutat* **10**(5): 400-407

Kukulski, W., Schenk, A. D., Johanson, U., Braun, T., de Groot, B. L., Fotiadis, D., Kjellbom, P. and Engel, A. (2005). The 5Å structure of heterologously expressed plant aquaporin SoPIP2;1. *J Mol Biol* **350**(4): 611-616

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**(5259): 680-685

Lagree, V., Froger, A., Deschamps, S., Hubert, J. F., Delamarche, C., Bonnec, G., Thomas, D., Gouranton, J. and Pellerin, I. (1999). Switch from an aquaporin to a glycerol channel by two amino acids substitution. *J Biol Chem* **274**(11): 6817-6819

Laize, V., Gobin, R., Rousselet, G., Badier, C., Hohmann, S., Ripoche, P. and Tacnet, F. (1999). Molecular and functional study of *AQY1* from *Saccharomyces cerevisiae*: role of the C-terminal domain. *Biochem Biophys Res Commun* **257**(1): 139-144

Laize, V., Tacnet, F., Ripoche, P. and Hohmann, S. (2000). Polymorphism of *Saccharomyces cerevisiae* aquaporins. *Yeast* **16**(10): 897-903

Lathrop, B. K., Burack, W. R., Biltonen, R. L. and S. Rulet, G. (1992). Expression of a group II phospholipase A2 from the venom of *Agkistrodon piscivorus piscivorus* in *Escherichia coli*: Recovery and renaturation from bacterial inclusion bodies. *Protein Expr Purifi* **3**(6): 512-517

Latijnhouwers, M. and Govers, F. (2003). A *Phytophthora infestans* G-protein β subunit is involved in sporangium formation. *Eukaryot Cell* **2**(5): 971-977

Latijnhouwers, M., Ligterink, W., Vleeshouwers, V. G., van West, P. and Govers, F. (2004). A Gα subunit controls zoospore motility and virulence in the potato late blight pathogen *Phytophthora infestans*. *Mol Microbiol* **51**(4): 925-936

Lee, J. K., Kozono, D., Remis, J., Kitagawa, Y., Agre, P. and Stroud, R. M. (2005). Structural basis for conductance by the archaeal aquaporin AqpM at 1.68 A. *Proc Natl Acad Sci U S A* **102**(52): 18932-18937

LeFurgey, A., Ingram, P. and Blum, J. J. (2001). Compartmental responses to acute osmotic stress in *Leishmania major* result in rapid loss of Na^+ and Cl^-. *Comp Biochem Physiol A Mol Integr Physiol* **128**(2): 385-394

Leigh, R. A., Pope, A. J., Jennings, I. R. and Sanders, D. (1992). Kinetics of the vacuolar H^+-pyrophosphatase: The roles of magnesium, pyrophosphate, and their complexes as substrates, activators, and inhibitors. *Plant Physiol.* **100**(4): 1698-1705

Lemercier, G., Dutoya, S., Luo, S., Ruiz, F. A., Rodrigues, C. O., Baltz, T., Docampo, R. and Bakalara, N. (2002). A vacuolar-type H^+-pyrophosphatase governs maintenance of functional acidocalcisomes and growth of the insect and mammalian forms of *Trypanosoma brucei*. *J Biol Chem* **277**(40): 37369-37376

Lerchl, J., Konig, S., Zrenner, R. and Sonnewald, U. (1995). Molecular cloning, characterization and expression analysis of isoforms encoding tonoplast-bound proton-translocating inorganic pyrophosphatase in tobacco. *Plant Mol Biol* **29**(4): 833-840

Letunic, I., Copley, R. R., Schmidt, S., Ciccarelli, F. D., Doerks, T., Schultz, J., Ponting, C. P. and Bork, P. (2004). SMART 4.0: towards genomic data integration. *Nucleic Acids Res* **32**: D142-144

Lin, H. H., Pan, Y. J., Hsu, S. H., Van, R. C., Hsiao, Y. Y., Chen, J. H. and Pan, R. L. (2005). Deletion mutation analysis on C-terminal domain of plant vacuolar H^+-pyrophosphatase. *Arch Biochem Biophys* **442**(2): 206-213

Liu, K., Nagase, H., Huang, C. G., Calamita, G. and Agre, P. (2006). Purification and functional characterization of aquaporin-8. *Biol Cell* **98**(3): 153-161

Liu, Z., Shen, J., Carbrey, J. M., Mukhopadhyay, R., Agre, P. and Rosen, B. P. (2002). Arsenite transport by mammalian aquaglyceroporins AQP7 and AQP9. *Proc Natl Acad Sci U S A* **99**(9): 6053-6058

Lodish, H. F. (2000). Molecular cell biology. New York, W.H. Freeman.

Lu, Q., Callahan, M., Hosfield, T., Neiditch, B., Greener, A., Bauer, J. C. and Kobrin, M. (1997). Generate high-quality, directional plasmid cDNA libraries. *Stratagene Newsletter* **10**(3): 119-120

Luo, S., Marchesini, N., Moreno, S. N. and Docampo, R. (1999). A plant-like vacuolar H^+-pyrophosphatase in *Plasmodium falciparum*. *FEBS Lett* **460**(2): 217-220

Maddy, A. H. (1976). A critical evaluation of the analysis of membrane proteins by polyacrylamide gel electrophoresis in the presence of dodecyl sulphate. *J Theor Biol* **62**(2): 315-326

Maeshima, M. (1991). H^+-translocating inorganic pyrophosphatase of plant vacuoles. Inhibition by Ca^{2+}, stabilization by Mg^{2+} and immunological comparison with other inorganic pyrophosphatases. *Eur J Biochem* **196**(1): 11-17

Maeshima, M. (1992). Characterization of the major integral protein of vacuolar membrane. *Plant Physiol* **98**(4): 1248-1254

Maeshima, M. (2000). Vacuolar H^+-pyrophosphatase. *Biochim Biophys Acta* **1465**(1-2): 37-51

Maeshima, M. (2001). Tonoplast transporters: Organization and function [Review]. *Ann Rev Plant Physiol Plant Mol Biol* **52**: 469-479

Maeshima, M. and Yoshida, S. (1989). Purification and properties of vacuolar membrane proton-translocating inorganic pyrophosphatase from mung bean. *J Biol Chem* **264**(33): 20068-20073

Malinen, A. M., Belogurov, G. A., Salminen, M., Baykov, A. A. and Lahti, R. (2004). Elucidating the role of conserved glutamates in H^+-pyrophosphatase of *Rhodospirillum rubrum*. *J Biol Chem* **279**(26): 26811-26816

Mansurova, S. E. (1989). Inorganic pyrophosphate in mitochondrial metabolism. *Biochim Biophys Acta* **977**(3): 237-247

Marchesini, N., Luo, S., Rodrigues, C. O., Moreno, S. N. and Docampo, R. (2000). Acidocalcisomes and a vacuolar H^+-pyrophosphatase in malaria parasites. *Biochem J* **347**: 243-253

Marchesini, N., Ruiz, F. A., Vieira, M. and Docampo, R. (2002). Acidocalcisomes are functionally linked to the contractile vacuole of *Dictyostelium discoideum*. *J Biol Chem* **277**(10): 8146-8153

Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWeese-Scott, C., Geer, L. Y., Gwadz, M., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Marchler, G. H., Mullokandov, M., Shoemaker, B. A., Simonyan, V., Song, J. S., Thiessen, P. A., Yamashita, R. A., Yin, J. J., Zhang, D. and Bryant, S. H. (2005). CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* **33**: D192-196

Markoff, A., Savov, A., Vladimirov, V., Bogdanova, N., Kremensky, I. and Ganev, V. (1997). Optimization of single-strand conformation polymorphism analysis in the presence of polyethylene glycol. *Clin Chem* **43**(1): 30-33

Martinez, R., Wang, Y., Benaim, G., Benchimol, M., de Souza, W., Scott, D. A. and Docampo, R. (2002). A proton pumping pyrophosphatase in the Golgi apparatus and plasma membrane vesicles of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **120**(2): 205-213

Martinoia, E., Massonneau, A. and Frangne, N. (2000). Transport processes of solutes across the vacuolar membrane of higher plants. *Plant Cell Physiol* **41**(11): 1175-1186

Maruyama, C., Tanaka, Y., Takeyasu, K., Yoshida, M. and Sato, M. H. (1998). Structural studies of the vacuolar H^+-pyrophosphatase: sequence analysis and identification of the residues modified by fluorescent cyclohexylcarbodiimide and maleimide. *Plant Cell Physiol* **39**(10): 1045-1053

Matsuzaki, T., Tajika, Y., Tserentsoodol, N., Suzuki, T., Aoki, T., Hagiwara, H. and Takata, K. (2002). Aquaporins: a water channel family. *Anat Sci Int* **77**(2): 85-93

Maurel, C. (1997). Aquaporins and water permeability of plant membranes. *Annu Rev Plant Physiol Plant Mol Biol* **48**: 399-429

Maurel, C. and Chrispeels, M. J. (2001). Aquaporins. A molecular entry into plant water relations. *Plant Physiol* **125**(1): 135-138

Maurel, C., Reizer, J., Schroeder, J. I., Chrispeels, M. J. and Saier, M. H., Jr. (1994). Functional characterization of the *Escherichia coli* glycerol facilitator, GlpF, in *Xenopus* oocytes. *J Biol Chem* **269**(16): 11869-11872

McCarren, K. L., McComb, J. A., Shearer, B. L. and Hardy, G. E. S. (2005). The role of chlamydospores of *Phytophthora cinnamomi* - a review. *Austr Plant Pathol* **34**(3): 333-338

McGuffin, L. J., Bryson, K. and Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**(4): 404-405

McIntosh, M. T., Drozdowicz, Y. M., Laroiya, K., Rea, P. A. and Vaidya, A. B. (2001). Two classes of plant-like vacuolar-type H^+-pyrophosphatases in malaria parasites. *Mol Biochem Parasitol* **114**(2): 183

McIntosh, M. T. and Vaidya, A. B. (2002). Vacuolar type H^+ pumping pyrophosphatases of parasitic protozoa. *Int J Parasitol* **32**(1): 1-14

McLeod, A., Smart, C. D. and Fry, W. E. (2004). Core promoter structure in the oomycete *Phytophthora infestans*. *Eukaryot Cell* **3**(1): 91-99

Meyrial, V., Laize, V., Gobin, R., Ripoche, P., Hohmann, S. and Tacnet, F. (2001). Existence of a tightly regulated water channel in *Saccharomyces cerevisiae*. *Eur J Biochem* **268**(2): 334-343

Mimotopes manual (2001). Antipeptide antibodies.

Mimura, H., Nakanishi, Y., Hirono, M. and Maeshima, M. (2004). Membrane topology of the H^+-pyrophosphatase of *Streptomyces coelicolor* determined by cysteine-scanning mutagenesis. *J Biol Chem* **279**(33): 35106-35112

Mimura, H., Nakanishi, Y. and Maeshima, M. (2005). Oligomerization of H^+-pyrophosphatase and its structural and functional consequences. *Biochim Biophys Acta* **1708**(3): 393-403

Mitchell, H. J. and Hardham, A. R. (1999). Characterisation of the water expulsion vacuole in *Phytophthora nicotianae* zoospores. *Protoplasma* **206**: 118-130

Mitra, B. N., Yoshino, R., Morio, T., Yokoyama, M., Maeda, M., Urushihara, H. and Tanaka, Y. (2000). Loss of a member of the aquaporin gene family, *aqpA* affects spore dormancy in *Dictyostelium*. *Gene* **251**(2): 131-139

Mitsuda, N., Enami, K., Nakata, M., Takeyasu, K. and Sato, M. H. (2001). Novel type *Arabidopsis thaliana* H^+-PPase is localized to the Golgi apparatus. *FEBS Lett* **488**(1-2): 29-33

Molecular Probes manual (2002). EnzChek Phosphatase Assay Kit Product Information Sheet (E-6646).

Montalvetti, A., Rohloff, P. and Docampo, R. (2004). A functional aquaporin co-localizes with the vacuolar proton pyrophosphatase to acidocalcisomes and the contractile vacuole complex of *Trypanosoma cruzi*. *J Biol Chem* **279**(37): 38673-38682

Montigny, C., Penin, F., Lethias, C. and Falson, P. (2004). Overcoming the toxicity of membrane peptide expression in bacteria by upstream insertion of Asp-Pro sequence. *Biochim Biophys Acta* **1660**(1-2): 53-65

Moraes Moreira, B. L., Soares Medeiros, L. C., Miranda, K., de Souza, W., Hentschel, J., Plattner, H. and Barrabin, H. (2005). Kinetics of pyrophosphate-driven proton uptake by acidocalcisomes of *Leptomonas wallacei*. *Biochem Biophys Res Commun* **334**(4): 1206-1213

Morishita, Y., Matsuzaki, T., Hara-chikuma, M., Andoo, A., Shimono, M., Matsuki, A., Kobayashi, K., Ikeda, M., Yamamoto, T., Verkman, A., Kusano, E., Ookawara, S., Takata, K., Sasaki, S. and Ishibashi, K. (2005). Disruption of aquaporin-11 produces polycystic kidneys following vacuolization of the proximal tubule. *Mol Cell Biol* **25**(17): 7770-7779

Morishita, Y., Sakube, Y., Sasaki, S. and Ishibashi, K. (2004). Molecular mechanisms and drug development in aquaporin water channel diseases: aquaporin superfamily (superaquaporins): Expansion of aquaporins restricted to multicellular organisms. *J Pharmacol Sci* **96**(3): 276-279

Morris, B. M. and Gow, N. A. R. (1993). Mechanism of electrotaxis of zoospores of phytopathogenic fungi. *Phytopathology* **83**(8): 877-882

Morris, P. F. and Ward, E. W. B. (1992). Chemoattraction of zoospores of the soybean pathogen, *Phytophthora sojae*, by isoflavones. *Physiol Mol Plant Pathol* **40**(1): 17-22

Motta, L. S., da Silva, W. S., Oliveira, D. M., de Souza, W. and Machado, E. A. (2004). A new model for proton pumping in animal cells: the role of pyrophosphate. *Insect Biochem Mol Biol* **34**(1): 19-27

Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J. B., Engel, A. and Fujiyoshi, Y. (2000). Structural determinants of water permeation through aquaporin-1. *Nature* **407**(6804): 599-605

Nakanishi, Y. and Maeshima, M. (1998). Molecular cloning of vacuolar H^+-pyrophosphatase and its developmental expression in growing hypocotyl of mung bean. *Plant Physiol* **116**(2): 589-597

Nakanishi, Y., Matsuda, N., Aizawa, K., Kashiyama, T., Yamamoto, K., Mimura, T., Ikeda, M. and Maeshima, M. (1999). Molecular cloning and sequencing of the cDNA for vacuolar H^+-pyrophosphatase from *Chara corallina*. *Biochim Biophys Acta* **1418**(1): 245-250

Nakanishi, Y., Saijo, T., Wada, Y. and Maeshima, M. (2001). Mutagenic analysis of functional residues in putative substrate-binding site and acidic domains of vacuolar H^+-pyrophosphatase. *J Biol Chem* **276**(10): 7654-7660

Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., McDonald, L., Utterback, T. R., Malek, J. A., Linher, K. D., Garrett, M. M., Stewart, A. M., Cotton, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton, G. G., Fleischmann, R. D., Eisen, J. A., White, O., Salzberg, S. L., Smith, H. O., Venter, J. C. and Fraser, C. M. (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**(6734): 323-329

Nelson, N. and Klionsky, D. J. (1996). Vacuolar H^+-ATPase: from mammals to yeast and back. *Experientia* **52**(12): 1101-1110

Niemietz, C. M. and Tyerman, S. D. (2002). New potent inhibitors of aquaporins: silver and gold compounds inhibit aquaporins of plant and human origin. *FEBS Lett* **531**(3): 443-447

Nolta, K. V. and Steck, T. L. (1994). Isolation and initial characterization of the bipartite contractile vacuole complex from *Dictyostelium discoideum*. *J Biol Chem* **269**(3): 2225-2233

Novagen manual (2004). pETBlue™ system manual.

Nyren, P., Nore, B. F. and Strid, A. (1991). Proton-pumping N,N'-dicyclohexylcarbodiimide-sensitive inorganic pyrophosphate synthase from *Rhodospirillum rubrum*: Purification, characterization, and reconstitution. *Biochemistry* **30**(11): 2883-2887

Olken, F. (2002). Phylogenetic tree computation tutorial.
(http://pga.lbl.gov/Workshop/April2002/lectures/Olken.pdf)

Orita, M., Suzuki, Y., Sekiya, T. and Hayashi, K. (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**(4): 874-879

Panabieres, F., Amselem, J., Galiana, E. and Le Berre, J.-Y. (2005). Gene identification in the oomycete pathogen *Phytophthora parasitica* during in vitro vegetative growth through expressed sequence tags. *Fungal Genet Biol* **42**(7): 611

Pasquier, C., Promponas, V. J., Palaios, G. A., Hamodrakas, J. S. and Hamodrakas, S. J. (1999). A novel method for predicting transmembrane segments in proteins based on a statistical analysis of the SwissProt database: the PRED-TMR algorithm. *Protein Eng* **12**(5): 381-385

Patterson, D. J. (1980). Contractile vacuoles and associated structures - their organization and function. *Biol Rev Camb Philos Soc* **55**(1): 1-46

Patterson, D. J. (1999). The diversity of eukaryotes. *Am Nat* **154**(S4): S96-S124

Pavlovic-Djuranovic, S., Schultz, J. E. and Beitz, E. (2003). A single aquaporin gene encodes a water/glycerol/urea facilitator in *Toxoplasma gondii* with similarity to plant tonoplast intrinsic proteins. *FEBS Lett* **555**(3): 500-504

Perez-Castineira, J. R., Lopez-Marques, R. L., Villalba, J. M., Losada, M. and Serrano, A. (2002). Functional complementation of yeast cytosolic pyrophosphatase by bacterial and plant H^+-translocating pyrophosphatases. *Proc Natl Acad Sci U S A* **99**(25): 15914-15919

Perkin-Elmer manual Antibodies: From design to assay - A practical guide.

Pettersson, N., Filipsson, C., Becit, E., Brive, L. and Hohmann, S. (2005). Aquaporins in yeasts and filamentous fungi. *Biol Cell* **97**(7): 487-500

'Pierce Chemical Technical Library' manual Antibody production.

Pieterse, C. M., van West, P., Verbakel, H. M., Brasse, P. W., van den Berg-Velthuis, G. C. and Govers, F. (1994). Structure and genomic organization of the ipiB and ipiO gene clusters of *Phytophthora infestans*. *Gene* **138**(1-2): 67-77

Poolman, B. and Glaasker, E. (1998). Regulation of compatible solute accumulation in bacteria. *Mol Microbiol* **29**(2): 397-407

Preston, G. M. and Agre, P. (1991). Isolation of the cDNA for erythrocyte integral membrane protein of 28 kilodaltons: member of an ancient channel family. *Proc Natl Acad Sci U S A* **88**(24): 11110-11114

Preston, G. M., Carroll, T. P., Guggino, W. B. and Agre, P. (1992). Appearance of water channels in Xenopus oocytes expressing red cell CHIP28 protein. *Science* **256**(5055): 385-387

Preston, G. M., Smith, B. L., Zeidel, M. L., Moulds, J. J. and Agre, P. (1994). Mutations in aquaporin-1 in phenotypically normal humans without functional CHIP water channels. *Science* **265**(5178): 1585-1587

Promponas, V. J., Palaios, G. A., Pasquier, C. M., Hamodrakas, J. S. and Hamodrakas, S. J. (1999). CoPreTHi: a Web tool which combines transmembrane protein segment prediction methods. *In Silico Biol* **1**(3): 159-162

Quigley, F., Rosenberg, J., Shachar-Hill, Y. and Bohnert, H. (2001). From genome to function: the *Arabidopsis* aquaporins. *Genome Biol* **3**(1): research0001.0001-0001.0017

Qutob, D., Hraber, P. T., Sobral, B. W. and Gijzen, M. (2000). Comparative analysis of expressed sequences in *Phytophthora sojae*. *Plant Physiol* **123**(1): 243-254

Randall, T. A., Dwyer, R. A., Huitema, E., Beyer, K., Cvitanich, C., Kelkar, H., Fong, A. M., Gates, K., Roberts, S., Yatzkan, E., Gaffney, T., Law, M., Testa, A., Torto-Alalibo, T., Zhang, M., Zheng, L., Mueller, E., Windass, J., Binder, A., Birch, P. R., Gisi, U., Govers, F., Gow, N. A., Mauch, F., van West, P., Waugh, M. E., Yu, J., Boller, T., Kamoun, S., Lam, S. T. and Judelson, H. S. (2005). Large-scale gene discovery in the oomycete *Phytophthora infestans* reveals likely components of phytopathogenicity shared with true fungi. *Mol Plant-Microbe Interact* **18**(3): 229-243

Randall, T. A. and Judelson, H. S. (1999). Construction of a bacterial artificial chromosome library of *Phytophthora infestans* and transformation of clones into *P. infestans*. *Fungal Genet Biol* **28**(3): 160-170

Rea, P. A., Britten, C. J., Jennings, I. R., Calvert, C. M., Skiera, L. A., Leigh, R. A. and Sanders, D. (1992). Regulation of vacuolar H+-pyrophosphatase by free calcium : A reaction kinetic analysis. *Plant Physiol* **100**(4): 1706-1715

Rea, P. A. and Poole, R. J. (1985). Proton-translocating inorganic pyrophosphatase in red beet (*Beta vulgaris* L.) tonoplast vesicles. *Plant Physiol* **77**(1): 46-52

Rea, P. A. and Poole, R. J. (1993). Vacuolar H^+ -translocating pyrophosphatase. *Ann Rev Plant Physiol Plant Mol Biol* **44**(1): 157-180

Ren, G., Reddy, V. S., Cheng, A., Melnyk, P. and Mitra, A. K. (2001). Visualization of a water-selective pore by electron crystallography in vitreous ice. *Proc Natl Acad Sci U S A* **98**(4): 1398-1403

Riethmuller, A., Voglmayr, H., Goker, M., Weiss, M. and Oberwinkler, F. (2002). Phylogenetic relationships of the downy mildews (Peronosporales) and related groups based on nuclear large subunit ribosomal DNA sequences. *Mycologia* **94**(5): 834-849

Ringler, P., Borgnia, M. J., Stahlberg, H., Maloney, P. C., Agre, P. and Engel, A. (1999). Structure of the water channel AqpZ from *Escherichia coli* revealed by electron crystallography. *J Mol Biol* **291**(5): 1181-1190

Rizzo, D. M., Garbelotto, M. and Hansen, E. M. (2005). *Phytophthora ramorum*: Integrative research and management of an emerging pathogen in California and Oregon forests. *Annu Rev Phytopathol* **43**: 309-335

Robinson, D. G., Haschke, H.-P., Hinz, G., Hoh, B., Maeshima, M. and Marty, F. (1996). Immunological detection of tonoplast polypeptides in the plasma membrane of pea cotyledons. *Planta* **198**(1): 95-103

Robinson, D. G., Hoppenrath, M., Oberbeck, K., Luykx, P. and Ratajczak, R. (1998). Localization of pyrophosphatase and V-ATPase in *Chlamydomonas reinhardtii*. *Bot Acta* **111**: 108-122

Robold, A. V. and Hardham, A. R. (1998). Production of species-specific monoclonal antibodies that react with surface components on zoospores and cysts of *Phytophthora nicotianae*. *Can J Microbiol* **44**(12): 1161-1170

Robold, A. V. and Hardham, A. R. (2004). Production of monoclonal antibodies against peripheral-vesicle proteins in zoospores of *Phytophthora nicotianae*. *Protoplasma* **223**(2 - 4): 121-132

Robold, A. V. and Hardham, A. R. (2005). During attachment *Phytophthora* spores secrete proteins containing thrombospondin type 1 repeats. *Curr Genet* **47**(5): 307-315

'Roche Applied Science' manual (2005). 5'/3' RACE Kit.

Rodrigues, C. O., Scott, D. A., Bailey, B. N., de Souza, W., Benchimol, M., Moreno, B., Urbina, J. A., Oldfield, E. and Moreno, S. N. J. (2000). Vacuolar proton pyrophosphatase activity and pyrophosphate (PP_i) in *Toxoplasma gondii* as possible chemotherapeutic targets. *Biochem J* **349**: 737-745

Rodrigues, C. O., Scott, D. A. and Docampo, R. (1999a). Characterization of a vacuolar pyrophosphatase in *Trypanosoma brucei* and its localization to acidocalcisomes. *Mol Cell Biol* **19**(11): 7712-7723

Rodrigues, C. O., Scott, D. A. and Docampo, R. (1999b). Presence of a vacuolar H^+-pyrophosphatase in promastigotes of *Leishmania donovani* and its localization to a different compartment from the vacuolar H^+-ATPase. *Biochem J* **340**: 759-766

Rodriguez, M. C., Froger, A., Rolland, J. P., Thomas, D., Aguero, J., Delamarche, C. and Garcia-Lobo, J. M. (2000). A functional water channel protein in the pathogenic bacterium *Brucella abortus*. *Microbiology* **146**: 3251-3257

Rohloff, P., Montalvetti, A. and Docampo, R. (2004). Acidocalcisomes and the contractile vacuole complex are involved in osmoregulation in *Trypanosoma cruzi*. *J Biol Chem* **279**(50): 52270-52281

Rost, B., Casadio, R., Fariselli, P. and Sander, C. (1995). Transmembrane helices predicted at 95% accuracy. *Protein Sci* **4**(3): 521-533

Rost, B., Sander, C. and Schneider, R. (1994). PHD--an automatic mail server for protein secondary structure prediction. *Comput Appl Biosci* **10**(1): 53-60

Roux, B. and Schulten, K. (2004). Computational studies of membrane channels. *Structure* **12**(8): 1343-1351

Ruiz, F. A., Marchesini, N., Seufferheld, M., Govindjee and Docampo, R. (2001a). The polyphosphate bodies of *Chlamydomonas reinhardtii* possess a proton-pumping pyrophosphatase and are similar to acidocalcisomes. *J Biol Chem* **276**(49): 46196-46203

Ruiz, F. A., Rodrigues, C. O. and Docampo, R. (2001b). Rapid changes in polyphosphate content within acidocalcisomes in response to cell growth, differentiation, and environmental stress in *Trypanosoma cruzi*. *J Biol Chem* **276**(28): 26114-26121

Sakakibara, Y., Kobayashi, H. and Kasamo, K. (1996). Isolation and characterization of cDNAs encoding vacuolar H^+-pyrophosphatase isoforms from rice (*Oryza sativa* L.). *Plant Mol Biol* **31**(5): 1029-1038

Sakurai, J., Ishikawa, F., Yamaguchi, T., Uemura, M. and Maeshima, M. (2005). Identification of 33 rice aquaporin genes and analysis of their expression and function. *Plant Cell Physiol* **46**(9): 1568-1577

Saliba, K. J., Allen, R. J., Zissis, S., Bray, P. G., Ward, S. A. and Kirk, K. (2003). Acidification of the malaria parasite's digestive vacuole by a H^+-ATPase and a H^+-pyrophosphatase. *J Biol Chem* **278**(8): 5605-5612

Sambrook, J. and Russell, D. W. (2001). Molecular cloning: A laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory.

Santoni, V., Gerbeau, P., Javot, H. and Maurel, C. (2000). The high diversity of aquaporins reveals novel facets of plant membrane functions. *Curr Opin Plant Biol* **3**(6): 476-481

Saparov, S. M., Kozono, D., Rothe, U., Agre, P. and Pohl, P. (2001). Water and ion permeation of aquaporin-1 in planar lipid bilayers. Major differences in structural determinants and stoichiometry. *J Biol Chem* **276**(34): 31515-31520

Sarafian, V., Kim, Y., Poole, R. J. and Rea, P. A. (1992). Molecular cloning and sequence of cDNA encoding the pyrophosphate-energized vacuolar membrane proton pump of *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **89**(5): 1775-1779

Sarafian, V. and Poole, R. J. (1989). Purification of an H^+-translocating inorganic pyrophosphatase from vacuole membranes of red beet. *Plant Physiol* **91**(1): 34-38

Sato, M. H., Maeshima, M., Ohsumi, Y. and Yoshida, M. (1991). Dimeric structure of H^+-translocating pyrophosphatase from pumpkin vacuolar membranes. *FEBS Lett* **290**(1-2): 177-180

Savage, D. F., Egea, P. F., Robles-Colmenares, Y., Iii, J. D. O., Connell and Stroud, R. M. (2003). Architecture and Selectivity in Aquaporins: 2.5 Å X-Ray Structure of Aquaporin Z. *PLoS Biology* **1**(3): e72

Schenk, A. D., Werten, P. J., Scheuring, S., de Groot, B. L., Muller, S. A., Stahlberg, H., Philippsen, A. and Engel, A. (2005). The 4.5 A structure of human AQP2. *J Mol Biol* **350**(2): 278-289

Schocke, L. and Schink, B. (1998). Membrane-bound proton-translocating pyrophosphatase of *Syntrophus gentianae*, a syntrophically benzoate-degrading fermenting bacterium. *Eur J Biochem* **256**(3): 589-594

Schultz, A. and Baltscheffsky, M. (2003). Properties of mutated *Rhodospirillum rubrum* H^+-pyrophosphatase expressed in *Escherichia coli*. *Biochim Biophys Acta* **1607**(2-3): 141-151

Schultz, A. and Baltscheffsky, M. (2004). Inhibition studies on *Rhodospirillum rubrum* H^+-pyrophosphatase expressed in *Escherichia coli*. *Biochim Biophys Acta* **1656**(2-3): 156-165

Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. and Bork, P. (2000). SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res* **28**(1): 231-234

Schuurmans, J. A., van Dongen, J. T., Rutjens, B. P., Boonman, A., Pieterse, C. M. and Borstlap, A. C. (2003). Members of the aquaporin family in the developing pea seed coat include representatives of the PIP, TIP, and NIP subfamilies. *Plant Mol Biol* **53**(5): 655–667

Scott, D. A., de Souza, W., Benchimol, M., Zhong, L., Lu, H. G., Moreno, S. N. and Docampo, R. (1998). Presence of a plant-like proton-pumping pyrophosphatase in acidocalcisomes of *Trypanosoma cruzi*. *J Biol Chem* **273**(34): 22151-22158

Seufferheld, M., Vieira, M. C., Ruiz, F. A., Rodrigues, C. O., Moreno, S. N. and Docampo, R. (2003). Identification of organelles in bacteria similar to acidocalcisomes of unicellular eukaryotes. *J Biol Chem* **278**(32): 29971-29978

Shan, W. and Hardham, A. R. (2004). Construction of a bacterial artificial chromosome library, determination of genome size, and characterization of an Hsp70 gene family in *Phytophthora nicotianae*. *Fungal Genet Biol* **41**(3): 369-380

Shan, W., Liu, J. and Hardham, A. R. (2006). *Phytophthora nicotianae* PnPMA1 encodes an atypical plasma membrane H^+-ATPase that is functional in yeast and developmentally regulated. *Fungal Genet Biol* **43**(8): 583-592

Shan, W., Marshall, J. S. and Hardham, A. R. (2004). Gene expression in germinated cysts of *Phytophthora nicotianae*. *Mol Plant Pathol* **5**(4): 317-330

Shiels, A. and Bassnett, S. (1996). Mutations in the founder of the MIP gene family underlie cataract development in the mouse. *Nature Genet* **12**(2): 212-215

Short, J. M., Fernandez, J. M., Sorge, J. A. and Huse, W. D. (1988). λ ZAP: a bacteriophage λ expression vector with *in vivo* excision properties. *Nucleic Acids Res* **16**(15): 7583-7600

Si-Ammour, A., Mauch-Mani, B. and Mauch, F. (2003). Quantification of induced resistance against *Phytophthora* species expressing GFP as a vital marker: β-aminobutyric acid but not BTH protects potato and *Arabidopsis* from infection. *Mol Plant Pathol* **4**(4): 237-248

Sidoux-Walter, F., Pettersson, N. and Hohmann, S. (2004). The *Saccharomyces cerevisiae* aquaporin Aqy1 is involved in sporulation. *Proc Natl Acad Sci U S A* **101**(50): 17422-17427

Škalamera, D., Wasson, A. P. and Hardham, A. R. (2004). Genes expressed in zoospores of *Phytophthora nicotianae*. *Mol Genet Genomics* **270**(6): 549-557

Smith, B. L. and Agre, P. (1991). Erythrocyte Mr 28,000 transmembrane protein exists as a multisubunit oligomer similar to channel proteins. *J Biol Chem* **266**(10): 6407-6415

Smith, B. L., Preston, G. M., Spring, F. A., Anstee, D. J. and Agre, P. (1994). Human red cell aquaporin CHIP. I. Molecular characterization of ABH and Colton blood group antigens. *J Clin Invest* **94**(3): 1043-1049

Soares Medeiros, L. C. A., Moreira, B. L. M., Miranda, K., de Souza, W., Plattner, H., Hentschel, J. and Barrabin, H. (2005). A proton pumping pyrophosphatase in acidocalcisomes of *Herpetomonas sp. Mol Biochem Parasitol* **140**(2): 175

Sogin, M. L. and Silberman, J. D. (1998). Evolution of the protists and protistan parasites from the perspective of molecular systematics. *Int J Parasitol* **28**(1): 11-20

Soltis, P. S. and Soltis, D. E. (2003). Applying the bootstrap in phylogeny reconstruction. *Stat Sci* **18**(2): 256-267

Steinfeld, S., Cogan, E., King, L. S., Agre, P., Kiss, R. and Delporte, C. (2001). Abnormal distribution of aquaporin-5 water channel protein in salivary glands from Sjogren's syndrome patients. *Lab Invest* **81**(2): 143-148

Stevens, T. H. and Forgac, M. (1997). Structure, function and regulation of the vacuolar H^+-ATPase. *Annu Rev Cell Dev Biol* **13**: 779-808

Stroud, R. M., Miercke, L. J., O'Connell, J., Khademi, S., Lee, J. K., Remis, J., Harries, W., Robles, Y. and Akhavan, D. (2003). Glycerol facilitator GlpF and the associated aquaporin family of channels. *Curr Opin Struct Biol* **13**(4): 424-431

Sui, H., Han, B. G., Lee, J. K., Walian, P. and Jap, B. K. (2001). Structural basis of water-specific transport through the AQP1 water channel. *Nature* **414**(6866): 872-878

Takasu, A., Nakanishi, Y., Yamauchi, T. and Maeshima, M. (1997). Analysis of the substrate binding site and carboxyl terminal region of vacuolar H^+-pyrophosphatase of mung bean with peptide antibodies. *J Biochem (Tokyo)* **122**(4): 883-889

Takata, K., Matsuzaki, T. and Tajika, Y. (2004). Aquaporins: water channel proteins of the cell membrane. *Prog Histochem Cytochem* **39**(1): 1-83

Tamas, M. J., Luyten, K., Sutherland, F. C., Hernandez, A., Albertyn, J., Valadi, H., Li, H., Prior, B. A., Kilian, S. G., Ramos, J., Gustafsson, L., Thevelein, J. M. and Hohmann, S. (1999). Fps1p controls the accumulation and release of the compatible solute glycerol in yeast osmoregulation. *Mol Microbiol* **31**(4): 1087-1104

Tanaka, Y., Chiba, K., Maeda, M. and Maeshima, M. (1993). Molecular cloning of cDNA for vacuolar membrane proton-translocating inorganic pyrophosphatase in *Hordeum vulgare*. *Biochem Biophys Res Commun* **190**(3): 1110

Tanghe, A., Van Dijck, P. and Thevelein, J. M. (2006). Why do microorganisms have aquaporins? *Trends Microbiol* **14**(2): 78-85

Tani, S. and Judelson, H. (2006). Activation of zoosporogenesis-specific genes in *Phytophthora infestans* involves a 7-nucleotide promoter motif and cold-induced membrane rigidity. *Eukaryot Cell* **5**(4): 745-752

Tani, S., Kim, K. S. and Judelson, H. S. (2005). A cluster of NIF transcriptional regulators with divergent patterns of spore-specific expression in *Phytophthora infestans*. *Fungal Genet Biol* **42**(1): 42-50

Temesvari, L. A., Rodriguez-Paris, J. M., Bush, J. M., Zhang, L. and Cardelli, J. A. (1996). Involvement of the vacuolar proton-translocating ATPase in multiple steps of the endo-lysosomal system and in the contractile vacuole system of *Dictyostelium discoideum*. *J Cell Sci* **109**: 1479-1495

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**(24): 4876-4882

Thornton, J. W. and DeSalle, R. (2000). Gene family evolution and homology: Genomics meets phylogenetics. *Annu Rev Genomics Hum Genet* **1**: 41-73

Tian, M., Benedetti, B. and Kamoun, S. (2005). A second Kazal-like protease inhibitor from *Phytophthora infestans* inhibits and interacts with the apoplastic pathogenesis-related protease P69B of tomato. *Plant Physiol* **138**(3): 1785-1793

Tian, M. and Kamoun, S. (2005). A two disulfide bridge Kazal domain from *Phytophthora* exhibits stable inhibitory activity against serine proteases of the subtilisin family. *BMC Biochem* **6**: 15

Tobias, J. W., Shrader, T. E., Rocap, G. and Varshavsky, A. (1991). The N-end rule in bacteria. *Science* **254**(5036): 1374-1377

Tooley, P. W. and Therrien, C. D. (1987). Cytophotometric determination of the nuclear-DNA content of 23 Mexican and 18 non-Mexican isolates of *Phytophthora infestans*. *Exp Mycol* **11**(1): 19-26

Tornroth-Horsefield, S., Wang, Y., Hedfalk, K., Johanson, U., Karlsson, M., Tajkhorshid, E., Neutze, R. and Kjellbom, P. (2006). Structural mechanism of plant aquaporin gating. *Nature* **439**(7077): 688-694

Tripathy, S., Pandey, V. N., Fang, B., Salas, F. and Tyler, B. M. (2006). VMD: A community annotation database for oomycetes and microbial genomes. *Nucleic Acids Res* **34**: D379-381

Tsubota, K., Hirai, S., King, L. S., Agre, P. and Ishida, N. (2001). Defective cellular trafficking of lacrimal gland aquaporin-5 in Sjogren's syndrome. *Lancet* **357**(9257): 688-689

Tusnady, G. E. and Simon, I. (1998). Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J Mol Biol* **283**(2): 489-506

Tusnady, G. E. and Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**(9): 849-850

Tyerman, S., Bohnert, H., Maurel, C., Steudle, E. and Smith, J. (1999). Plant aquaporins: their molecular biology, biophysics and significance for plant water relations. *J Exp Bot* **50**(90001): 1055-1071

Tyerman, S. D., Niemietz, C. M. and Bramley, H. (2002). Plant aquaporins: multifunctional water and solute channels with expanding roles. *Plant Cell Environ* **25**(2): 173-194

Tzeng, C. M., Yang, C. Y., Yang, S. J., Jiang, S. S., Kuo, S. Y., Hung, S. H., Ma, J. T. and Pan, R. L. (1996). Subunit structure of vacuolar proton-pyrophosphatase as determined by radiation inactivation. *Biochem J* **316**: 143-147

Unger, V. M. (2000). Fraternal twins: AQP1 and GlpF. *Nature Struct Biol* **7**(12): 1082-1084

Uzcategui, N. L., Szallies, A., Pavlovic-Djuranovic, S., Palmada, M., Figarella, K., Boehmer, C., Lang, F., Beitz, E. and Duszenko, M. (2004). Cloning, heterologous expression, and characterization of three aquaglyceroporins from *Trypanosoma brucei*. *J Biol Chem* **279**(41): 42669-42676

Van der Auwera, G. and De Wachter, R. (1996). Large-subunit rRNA sequence of the chytridiomycete *Blastocladiella emersonii*, and implications for the evolution of zoosporic fungi. *J Mol Evol* **43**(5): 476-483

Van, R. C., Pan, Y. J., Hsu, S. H., Huang, Y. T., Hsiao, Y. Y. and Pan, R. L. (2005). Role of transmembrane segment 5 of the plant vacuolar H^+-pyrophosphatase. *Biochim Biophys Acta* **1709**(1): 84-94

van West, P., Appiah, A. A. and Gow, N. A. R. (2003). Advances in research on oomycete root pathogens. *Physiol Mol Plant Pathol* **62**(2): 99-113

van West, P., de Jong, A. J., Judelson, H. S., Emons, A. M. and Govers, F. (1998). The *ipiO* gene of *Phytophthora infestans* is highly expressed in invading hyphae during infection. *Fungal Genet Biol* **23**(2): 126-138

van West, P., Kamoun, S., van 't Klooster, J. W. and Govers, F. (1999a). Internuclear gene silencing in *Phytophthora infestans*. *Mol Cell* **3**(3): 339-348

van West, P., Morris, B. M., Reid, B., Appiah, A. A., Osborne, M. C., Campbell, T. A., Shepherd, S. J. and Gow, N. A. R. (2002). Oomycete plant pathogens use electric fields to target roots. *Mol Plant-Microbe Interact* **15**(8): 790-798

van West, P., Reid, B., Campbell, T. A., Sandrock, R. W., Fry, W. E., Kamoun, S. and Gow, N. A. (1999b). Green fluorescent protein (GFP) as a reporter gene for the plant pathogenic oomycete *Phytophthora palmivora*. *FEMS Microbiol Lett* **178**(1): 71-80

Vidossich, P., Cascella, M. and Carloni, P. (2004). Dynamics and energetics of water permeation through the aquaporin channel. *Proteins* **55**(4): 924-931

Vijn, I. and Govers, F. (2003). *Agrobacterium tumefaciens* mediated transformation of the oomycete plant pathogen *Phytophthora infestans*. *Mol Plant Pathol* **4**(6): 459-467

Walker, J. M. (2005). The proteomics protocols handbook. Totowa, N.J., Humana Press.

Walker, R. R. and Leigh, R. A. (1981). Mg^{2+}-Dependent, cation-stimulated inorganic pyrophosphatase associated with vacuoles isolated from storage roots of red beet (*Beta vulgaris* L.). *Planta* **153**(2): 150-155

Walz, T., Hirai, T., Murata, K., Heymann, J. B., Mitsuoka, K., Fujiyoshi, Y., Smith, B. L., Agre, P. and Engel, A. (1997). The three-dimensional structure of aquaporin-1. *Nature* **387**(6633): 624-627

Wang, D. N., Safferling, M., Lemieux, M. J., Griffith, H., Chen, Y. and Li, X. D. (2003). Practical aspects of overexpressing bacterial secondary membrane transporters for structural studies. *Biochim Biophys Acta* **1610**(1): 23-36

Wang, Y., Schulten, K. and Tajkhorshid, E. (2005). What makes an aquaporin a glycerol channel? A comparative study of AqpZ and GlpF. *Structure (Camb)* **13**(8): 1107-1118

Washington, W. S. and McGee, P. (2000). Dimethomorph soil and seed treatment of potted tomatoes for control of damping-off and root rot caused by *Phytophthora nicotianae* var. *nicotianae*. *Austr Plant Pathol* **29**(1): 46-51

Waugh, M., Hraber, P., Weller, J., Wu, Y., Chen, G., Inman, J., Kiphart, D. and Sobral, B. (2000). The *Phytophthora* genome initiative database: informatics and analysis for distributed pathogenomic research. *Nucleic Acids Res* **28**(1): 87-90

Weiner, H., Stitt, M. and Heldt, H. W. (1987). Subcellular compartmentation of pyrophosphate and alkaline pyrophosphatase in leaves. *Biochim Biophys Acta* **893**(1): 13-21

Whisson, S. C., Avrova, A. O., Van West, P. and Jones, J. T. (2005). A method for double-stranded RNA-mediated transient gene silencing in *Phytophthora infestans*. *Mol Plant Pathol* **6**(2): 153-163

Whisson, S. C., van der Lee, T., Bryan, G. J., Waugh, R., Govers, F. and Birch, P. R. (2001). Physical mapping across an avirulence locus of *Phytophthora infestans* using a highly representative, large-insert bacterial artificial chromosome library. *Mol Genet Genomics* **266**(2): 289-295

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D. and Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **112**: 531-552

Win, J., Kanneganti, T. D., Torto-Alalibo, T. and Kamoun, S. (2006). Computational and comparative analyses of 150 full-length cDNA sequences from the oomycete plant pathogen *Phytophthora infestans*. *Fungal Genet Biol* **43**(1): 20-33

Winstanley, M., Woodward, S. and Walker, N. (1995). Tobacco in Australia: Facts and issues. Australia Victorian Smoking and Health Program (Quit Victoria).

Wu, M., Ren, Q., Durkin, A. S., Daugherty, S. C., Brinkac, L. M., Dodson, R. J., Madupu, R., Sullivan, S. A., Kolonay, J. F., Nelson, W. C., Tallon, L. J., Jones, K. M., Ulrich, L. E., Gonzalez, J. M., Zhulin, I. B., Robb, F. T. and Eisen, J. A. (2005). Life in hot carbon monoxide: The complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. *PLoS Genet* **1**(5): e65

Yancey, P. H., Clark, M. E., Hand, S. C., Bowlus, R. D. and Somero, G. N. (1982). Living with water stress: Evolution of osmolyte systems. *Science* **217**(4566): 1214-1222

Yang, S. J., Jiang, S. S., Hsiao, Y. Y., Van, R. C., Pan, Y. J. and Pan, R. L. (2004). Thermoinactivation analysis of vacuolar H^+-pyrophosphatase. *Biochim Biophys Acta* **1656**(2-3): 88-95

Yang, S. J., Jiang, S. S., Kuo, S. Y., Hung, S. H., Tam, M. F. and Pan, R. L. (1999). Localization of a carboxylic residue possibly involved in the inhibition of vacuolar H^+-pyrophosphatase by *N, N'*-dicyclohexylcarbodi-imide. *Biochem J* **342**: 641-646

Yang, S. J., Jiang, S. S., Van, R. C., Hsiao, Y. Y. and Pan, R.-L. (2000). A lysine residue involved in the inhibition of vacuolar H^+-pyrophosphatase by fluorescein 5'-isothiocyanate. *Biochim Biophys Acta Bioenerg* **1460**(2-3): 375-383

Yasui, M., Hazama, A., Kwon, T. H., Nielsen, S., Guggino, W. B. and Agre, P. (1999). Rapid gating and anion permeability of an intracellular aquaporin. *Nature* **402**(6758): 184-187

Yool, A. J. and Weinstein, A. M. (2002). New roles for old holes: ion channel function in aquaporin-1. *News Physiol Sci* **17**: 68-72

Zancani, M., Macri, F., Dal Belin Peruffo, A. and Vianello, A. (1995). Isolation of the catalytic subunit of a membrane-bound H^+-pyrophosphatase from pea stem mitochondria. *Eur J Biochem* **228**(1): 138-143

Zardoya, R. (2005). Phylogeny and evolution of the major intrinsic protein family. *Biol Cell* **97**(6): 397-414

Zhen, R. G., Baykov, A. A., Bakuleva, N. P. and Rea, P. A. (1994a). Aminomethylenediphosphonate: A potent type-specific inhibitor of both plant and phototrophic bacterial H+-pyrophosphatases. *Plant Physiol* **104**(1): 153-159

Zhen, R. G., Kim, E. J. and Rea, P. A. (1994b). Localization of cytosolically oriented maleimide-reactive domain of vacuolar H^+-pyrophosphatase. *J Biol Chem* **269**(37): 23342-23350

Zhen, R. G., Kim, E. J. and Rea, P. A. (1997). Acidic residues necessary for pyrophosphate-energized pumping and inhibition of the vacuolar H^+-pyrophosphatase by *N,N'*-dicyclohexylcarbodiimide. *J Biol Chem* **272**(35): 22340-22348

Appendix

Appendix A Buffers and solutions

The buffers and solutions used in this thesis research are described below. All buffers and solutions were stored at RT unless otherwise indicated. All percentages are weight per volume (w/v) unless otherwise stated.

A.1 DNA buffers and solutions

1 M Tris-HCl, pH 8: 121.1 g Tris(hydroxylmethyl)aminomethane was dissolved in 800 mL ddH_2O, adjusted to pH 8 by the addition of HCl and then made up to 1 L with ddH_2O. The buffer was sterilised by autoclaving at 121°C for 20 min.

TE buffer, pH 8: 10 mM Tris-HCl (pH 8), 1 mM Na_2EDTA (pH 8). The buffer was sterilised by autoclaving at 121°C for 20 min.

0.5 M Na_2EDTA, pH 8: 186.1 g Na_2EDTA was added to 800 mL ddH_2O. The pH was adjusted to 8 with NaOH, made up to 1 L with ddH_2O and sterilised by autoclaving at 121°C for 20 min.

5 M NaCl: 292.2 g NaCl was dissolved in 800 mL ddH_2O and made up to 1 L with ddH_2O. It was sterilised by autoclaving at 121°C for 20 min.

1 M $MgCl_2$: 219.1 g $CaCl_2.6H_2O$ was dissolved in 800mL ddH_2O and made up to 1 L with ddH_2O. It was sterilised by autoclaving at 121°C for 20 min.

1 M $CaCl_2$: 203.31 g $MgCl_2.6H_2O$ was dissolved in 800 mL ddH_2O and made up to 1 L with ddH_2O. It was sterilised by autoclaving at 121°C for 20 min.

1 M $MgSO_4$: 246 g $MgSO_4.7H_2O$ was dissolved in 800 mL ddH_2O and made up to 1 L with ddH_2O. It was sterilised by autoclaving at 121°C for 20 min.

0.1 M Phosphate buffer: Stock solutions: 27.2 g KH_2PO_4 was dissolved in 1 L dH_2O. 34.8 g K_2HPO_4 was dissolved in 1 L ddH_2O. For making 0.1 M buffer: 39 mL of KH_2PO_4 and 61 mL of K_2HPO_4 were mixed in a clean Schott bottle and the final pH checked.

10 M NaOH: 40 g of NaOH was dissolved in 100 mL of ddH_2O and sterilised by autoclaving at 121°C for 20 min.

3 M sodium acetate: 40.82 g of $NaOAc.3H_2O$ was dissolved in 80 mL sterile ddH_2O. The buffer was adjusted to pH 5.2 with glacial acetic acid, and made up to 100 mL and sterilised by autoclaving at 121°C for 20 min.

50x TAE: 242 g Tris-base plus 57.1 mL glacial acetic acid and 100 mL 0.5M Na_2EDTA (pH 8) were dissolved in ddH_2O to make 1 L. The buffer was sterilised by autoclaving at 121°C for 20 min. A 1x TAE working solution was used for nucleic acid gel electrophoresis.

6x DNA gel loading buffer: 0.025g bromophenol blue plus 0.025g xylene cyanol FF and 7 mL ddH_2O and 3 mL 100% glycerol were mixed together to make 10 mL.

Invitrogen 1 kb DNA Ladder (100 ng/µL): 100 µL of 1 µg/µL 1 kb DNA Ladder was mixed with 100 µL of DNA gel loading buffer and 800 µL 1x TAE buffer to a final volume of 1 mL. Stored at 4°C until use.

NEB 1 kb DNA Ladder (100 ng/µL): 200 µL of 0.5 µg/µL 1 kb DNA Ladder was mixed with 100 µL of DNA gel loading buffer and 700 µL 1x TAE buffer to a final volume of 1 mL. Stored at 4°C until use.

Ethidium bromide: 1 g of EtBr was stirred in 100 mL ddH_2O and stored in the dark. 20 µL/100 mL ddH_2O was used for staining DNA in agarose gels.

Solution I/Resuspension buffer - BAC and plasmid miniprep (100 mL)

50 mM glucose	5 mL of 1M glucose
10 mM EDTA	2 mL of 0.5 EDTA, pH 8
25 mM Tris-HCl	2.5 mL of 1 M Tris-HCl, pH 8
ddH_2O	90.5 mL

Solution II/Lysis buffer – BAC and plasmid miniprep (50 mL)

0.2 M NaOH	1 mL 10M NaOH
1% SDS	2.5 mL 20% SDS
ddH_2O	46.5 mL

Solution III/Neutralisation buffer - BAC/plasmid miniprep (100 mL):

3 M KOAc	60 mL 5 M KOAc
glacial acetic acid	11.5 mL
ddH_2O	28.5 mL

Phytophthora genomic DNA extraction buffer (100 mL):

7 M Urea	42 g
350 mM NaCl	2.05 g
4 mM EDTA	4 mL of 0.5M EDTA, pH 8
50 mM Tris-HCl, pH 8	5 mL of 1 M Tris-HCl, pH 8
1% sodium sarkosyl	1 g
5% phenol in 10 mM Tris-HCl, pH7.5	5 mL
dH_2O	85 mL

The buffer was sterilised by autoclaving at 121°C for 20 min.

A.2 RNA and SSCP buffers and solutions

DEPC treated water: 1 mL of DEPC was added to 1 L of sterile ddH_2O to inhibit RNAse activity, left overnight at 37°C and autoclaved the next day at 121°C for 20 min. All RNA solutions were made in DEPC treated H_2O. All glassware for RNA use was baked at 180°C for 4 h.

10x MOPS running buffer: 0.4 M MOPS, pH 7, 0.1 M sodium acetate; 0.01 M EDTA. The buffer was stored at 4°C for 3 months.

RNA gel loading buffer: 1 mM EDTA pH 8, 0.25% bromophenol blue, 0.25% xylene cyanol, 50% glycerol, 0.5 µg/mL EtBr made up to 10 mL and stored at –20°C until use.

SSCP - Running buffer (10x TBE): Mixed 108 g Tris (0.89 M), 55 g boric acid (0.89 M) and 40 mL 0.5M EDTA, pH 8.0 (20 mM) made up to 1 L, and stored at RT.

SSCP - Sample buffer: 1 mL was made up by adding 950 µL formamide, 5 µL of 10% solution of xylene cyanol and bromophenol blue, and 40 µL of 0.5 M EDTA, pH 8.0. The buffer was stored at -20°C.

SSCP gel (for one 28:1 (bis/acrylamide), 10% gel)

30% acrylamide/bis (37.5:1)	830 µL
30% acrylamide/bis (19:1)	830 µL
10x TBE, pH 8.0	250 µL
50% glycerol	500 µL
dH_2O	2.59 mL
10% Ammonium persulphate	40 µL
TEMED	5 µL

Silver staining – Fixer solution: 7.5 mL glacial acetic acid mixed with 92.5 mL ddH_2O. Stored at RT or 4°C.

Silver staining – Silver Nitrate solution: 150 mg $AgNO_3$ and 0.056% formaldehyde were dissolved in 100 mL ddH_2O approximately 15 min before use.

Silver staining – Developer solution: Mixed 300 mg Na_2CO_3, 0.056% formaldehyde and 40 µg sodium thiosulphate in 100 mL ddH_2O. Made up fresh before and stored at 8°C before use.

A.3 DNA/RNA blots transfer solutions and hybridisation buffers

20x SSC: Mixed 175.4 g NaCl with 88.2 g of sodium citrate dihydrate and brought to a final volume of 1 L in dH_2O and sterilised by autoclaving at 121°C for 20 min.

Depurination buffer: 11 mL HCl and 989 mL ddH_2O. Mixed and stored for up to a month.

Denaturation buffer: Mixed 87.66 g NaCl (1.5 M) and 20 g NaOH (0.5 M) in 1 L ddH_2O, stirred to dissolve and stored for up to a month.

Neutralisation buffer: Mixed 87.66 g NaCl (1.5 M) and 60.5 g Tris (0.5 M) in 1 L ddH_2O, stirred to dissolve. Adjusted pH to 7.5 with HCl and stored for up to a month.

50x Denhardts solution: Mixed 4.0 g BSA, 4.0 g Ficoll400, 4.0 g polyvinylpyrrolidone in 400 mL ddH_2O. Mixed to dissolve and stored at –20°C for up to 3 months.

20% SDS: Mixed 200 g SDS in 1 L sterile ddH_2O, heated to 70°C to solubilise the SDS.

Herring sperm DNA: Made to a final concentration of 10 mg/mL in sterile ddH_2O and stored at –20°C for up to 3 months.

Pre-hybridisation/Hybridisation buffer (10 mL):

6x SSC	3 mL of 20x SSC
5x Denhardts	1 mL of 50x Denhardts
0.5% SDS	0.25 mL of 20% SDS
100 µg/mL herring sperm DNA*	0.1 mL of 10 mg/mL herring sperm DNA
sterile ddH_2O	5.65 mL

* Boil for 10 min to denature DNA before adding to the buffer.

A.4 DIG Southern blot solutions

DIG Southern blot - Pre-hybridisation/Hybridisation buffer, pH 7.2 (200 mL)

0.25 M Na_2HPO_4	7.09g
1 mM EDTA	0.4 mL (0.5M EDTA solution)
20% SDS	40 g
0.5% Blocking Reagent (Roche)	10 mL 10x stock solution

DIG Southern blot - Washing buffer (500 mL)

20m M Na_2HPO_4	1.42 g
1 mM EDTA	1 mL 0.5 M EDTA solution
1% SDS	5 g

10x Maleic acid buffer, pH 7.5: 116.08 g Maleic acid (1 M) and 87.65 g NaCl (1.5 M) were added to 1 l ddH_2O. The buffer was sterilised by autoclaving at 121°C for 20 min and stored at RT.

DIG detection - Washing buffer, pH 8 (500 mL)

0.1 M Maleic acid	50 mL 10x Maleic acid buffer
3 M NaCl	83.27 g
0.3% Tween 20	1.5 mL

DIG detection - Blocking buffer: 5 mL 10x Blocking Reagent was diluted in 45 mL DIG Washing buffer prior to use.

Detection buffer, pH 9.5: 6.05 g Tris (0.1 M) and 2.92 g NaCl (0.1 M) were dissolved in 500 mL ddH_2O. The buffer was adjusted to pH 9.5 with HCl and sterilised by autoclaving at 121°C for 20 min.

A.5 Protein buffers and solutions

20x PBS Buffer: 58 g $Na_2HPO_4.12H_2O$ (80 mM), 4 g KCl (26 mM), 4 g KH_2PO_4 (15 mM) and 160 g NaCl (1.37 M) were dissolved in 2 L ddH_2O and stored at RT until use.

PBST: 1x PBS containing 0.05% Tween20. Stored at RT until use.

10x TBS: 31 g Tris (0.2 M) and 87 g NaCl (1.5 M) were dissolved in 1 L ddH_2O and the pH was adjusted to 7.5 with HCl. The buffer was store at RT.

TBST: 1x TBS containing 0.05% Tween20. Stored at RT until use.

500 mM PIPES, pH 7: 15.12g PIPES added to 100 mL ddH_2O. NaOH pellets were added until PIPES had dissolved and pH was 8. Stored at 4°C until use.

1 M HEPES: 23.8 g HEPES (N-(2-hydroxyethyl) piperazine-N'-(2-ethanesulphonic acid) were dissolved in 100 mL ddH_2O, sterilised by autoclaving at 121°C for 20 min and stored at RT.

100 mM EGTA: 38 g EDTA were added to 100 mL ddH_2O and stirred until dissolved. .The buffer was sterilised by autoclaving at 121°C for 20 min and stored at RT.

0.8 M MES-KOH, pH 6.5: 17 g MES (2-(N-morpholino)ethanesulphonic acid) were dissolved in 100 mL ddH_2O and the pH was adjusted to 6.5 with KOH. The buffer was stored at RT.

0.5 M Tris-MES buffer, pH 7.0: 7.88 g Tris were added to 100 mL ddH2O and the ph was adjusted to 7 with a 0.5 M MES solution. The buffer was stored at RT.

0.1 M glycine, pH 2.7 (Elution buffer): 7.5 g glycine was dissolved in 100 mL ddH_2O. The pH was adjusted to 2.7 with HCl and the buffer was store at RT.

2D-Rehydration buffer: 15.02 g urea (5 M) and 7.61 g thiourea (2 M) were dissolved in 25 mL ddH_2O. 2 g CHAPS (4%), 3-10 Pharmolyte (0.2%),4 mL 0.5 M Tris-HCl pH 6.8 (40 mM) and 0.5 g DTT (65 mM) were added and volume was made up to 50 mL with ddH_2O. The buffer was stored in aliquots at -20°C.

10x SDS-PAGE running buffer: 60.6 g Tris (25 mM), 288.2 g glycine (200 mM), 20 g SDS (1%) added to 2 L ddH_2O and stored at 4°C until use.

Coomassie staining solution: 50 mg Coomassie brilliant blue R-250 and 40 mL ethanol were mixed together, then 10 mL of glacial acetic acid and 50 mL of ddH_2O were added to the mixture. The staining solution was filtered and stored at 4°C until use.

De-staining solution: 40 mL ethanol, 10 mL glacial acetic acid and 50 mL of ddH_2O were mixed together and stored at 4°C until use.

Detection buffer, pH 9.5: 6.05 g Tris-HCl (0.1 M) and 2.92 g NaCl (0.1 M) were dissolved in 500 mL ddH_2O and stored at RT.

2x Sample buffer: 2 mL glycerol (10%), 1 mL β-mercaptoethanol (5%), 4 g SDS (2%), 2.5 mL 0.5 M Tris-HCl, pH 6,75, mL and 0.008 g bromophenol blue were made up to 20 ml with ddH_2O and stored in aliquots at -20°C until use.

SDS-PAGE - Resolving gel (for one 10% gel)

1.5 M Tris-HCl, pH 8.8	1.3 mL
10% SDS	50 µL
30% acrylamide/bis	1.7 mL
ddH2O	1.9 mL
TEMED	5 µL
Ammonium persulphate	50 µL

SDS-PAGE - Stacking gel (for one 5% gel)

0.5 M Tris-HCl, pH 6.8	0.5 mL
10% SDS	20 µL
30% acrylamide/bis	0.33 mL
ddH_2O	1.15 mL
TEMED	2 µL
Ammonium persulphate	20 µL

Immunofluorescence mounting medium: 5 mL Mowiol (10%), 15 mL glycerol (30%) and 10 mL 1 M Tris, pH8.5 (0.2 M) were made up to 50 mL with ddH_2O. The medium was stored in 1 mL aliquots at -20°C.

Microsomal preparation - 10x Osmoticum buffer: 72.85 g sorbitol (0.2 M), 2 g PVP-40 (0.1% w/v), 4 mL 1M $MgSO_4$ (2 mM), 2 mL 4 M KCl (5 mM), 12 mL 0.8 M MES-KOH (50 mM) and 0.4 ml 0.5 M EDTA (1 mM) were made up to 200 mL with ddH_2O. The buffer was stored in 25 mL aliquots at -20°C until use.

Microsomal preparation - Resupension Buffer: 1.7 g sucrose, 0.2 mL 0.5 M Tris-MES buffer (5 mM), 0.4 mL 0.1 M DTT solution (2 mM), 2 mL glycerol (10%) and 1 mM phenylmethyl sulphonyl fluoride were made up to 20 mL ddH_2O. The buffer was stored in 1 mL aliquots at -20°C until use.

V-PPase activity assay – 20x Reaction buffer (20 mL)

2.6 M KCl	3.876 g
0.2 M HEPES, pH7.2	4 mL of 1 M HEPES solution
40 mM MgSO4	8 mL of 100 mM $MgSO_4$ solution
1 mM EGTA	0.2 mL of 100 mM EGTA

Appendix B Bacterial strains and media

The bacterial media used in this thesis research are described below. All solutions were stored at RT unless otherwise indicated. All percentages are weight per volume (w/v) unless otherwise stated.

B.1 Bacterial strains

BACTERIAL STRAINS	GENOTYPE	SUPPLIER
E. coli DH10β	F' *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) *φ80lacZ*ΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ⁻ *rps*L *nup*G	Invitrogen
E coli XL1-Blue MRF′	Δ(*mcrA*)*183* Δ(*mcr*CB-*hsd*SMR-*mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [F′ *proAB lacI*^{q}ZΔM15 *Tn*10 (Tet^{R})]	Stratagene
E. coli SOLR™	e14⁻(McrA⁻) Δ(*mcr*CB-*hsd*SMR-*mrr*)171 *sbcC recB recJ uvrC umuC*::Tn5 (Kan^{R}) *lac gyrA96 relA1 thi-1 endA1* λ^{R} [F′ *proAB lacI*^{q}ZΔM15] *Su*⁻ (nonsuppressing)	Stratagene
E. coli Tuner™(DE3)pLacI	F⁻ *omp*T *hsd*SB(*rB*⁻ *mB*⁻) *gal dcm lac*Y1 (*DE3*) *pLacI* (Cam^{R})	Novagen

B.2 Bacterial media and solutions

IPTG (1 M): Dissolved 2.38 g IPTG in 10 mL sterile ddH_2O and stored in 1 mL aliquots at –20°C until use.

X-Gal (50 mg/mL): Dissolved 0.5 g X-Gal in 10 mL N, N' –dimethyl formamide and stored in 1 mL aliquots at –20°C until use.

LB Broth: 10 g NaCl, 10 g tryptone and 5 g yeast extract dissolved in 800 mL ddH_2O and after adjusting to pH 7 with 5 N NaOH, made up to 1 L and sterilised by autoclaving at 121°C for 20 min.

LB Broth + 0.2% maltose + $MgSO_4$: To 494 mL LB broth, added 1 g maltose + 5 mL 1M $MgSO_4$ and sterilised by filtering. .

LB Agar: 10 g NaCl, 10 g tryptone, 5 g yeast extract and 20 g agar were dissolved in 800 mL ddH_2O and after adjusting to pH 7 with 5 N NaOH, made up to 1 L and sterilised by autoclaving at 121°C for 20 min. Poured into Petri dishes (~25 mL/90 mm plate) and stored at 4°C until use.

LB Agar Top Agar: 10 g NaCl, 10 g tryptone, 5 g yeast extract and 7 g agar were dissolved in 800 mL ddH_2O and after adjusting to pH 7 with 5 N NaOH, made up to 1 L and sterilised by autoclaving at 121°C for 20 min.

NZY agar: 5 g NaCl, 5 g yeast extract, 15 g agar, 2 g $MgSO_4.7H_2O$, 10 g NZ amine (casein hydrolysate) were dissolved in 800 mL ddH_2O and after adjusting to pH 7 with 5 N NaOH, made up to 1 L and sterilised by autoclaving at 121°C for 20 min. Poured into Petri dishes (~25 mL/ 90 mm plate) and stored at 4°C until use.

NZY top agar: 5 g NaCl, 5 g yeast extract, 7 g agar (0.7%), 2 g $MgSO_4.7H_2O$, 10 g NZ amine were dissolved in 800 mL ddH_2O and after adjusting to pH 7 with 5 N NaOH, made up to 1 L and sterilised by autoclaving at 121°C for 20 min.

SOB medium: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl and 2.5 mL KCl were dissolved in 900 mL ddH_2O and after adjusting to pH 7 with 5 N NaOH, made up to 990 mL and sterilised by autoclaving at 121°C for 20 min. Before use added 10 mL of sterile 1 M $MgCl_2$.

SOC medium: Identical to SOB medium, except that it additionally contained 1 M glucose.

SM Buffer: 5.8 g NaCl, 2 g $MgSO_4.7H_2O$, 50 mL 1 M Tris-HCl (pH 7.5) and 5 mL of 2% gelatine were mixed in 1 L of ddH_2O and sterilised by autoclaving at 121°C for 20 min.

Colony wash solution (1 L)

50 mM Tris-HCl, pH8	50 mL of 1M Tris-HCl, pH 8
1 M NaCl	58.44 g
1 mM EDTA	2 mL of 0.5 EDTA, pH 8
0.1% SDS	5 mL of 20% SDS
ddH_2O	943 mL

TNT buffer, pH 7.5: 1.58 g Tris (0.1 M) and 0.88 g NaCl (0.15 M) were added to 100 mL ddH_2O and the pH was adjusted to 7.5 with HCl. 50 µL Tween20 (0.05%) was added and the buffer was stored at RT.

Appendix C Antibodies

ANTIBODY	DESCRIPTION	RAISED IN	SUPPLIER
SAM-FITC	Anti mouse F(ab')2 Fragment FITC-conjugated	Sheep	Chemicon
SAM-AP	Anti mouse Ig, AP-conjugated	Sheep	Silenus
Anti-mouse biotin	Anti mouse Ig, biotinylated	Donkey	GE Healthcare
SARa-FITC	Anti rabbit F(ab')2 Fragment FITC conjugated	Sheep	Chemicon
SARa-AP	Anti rabbit Ig, AP-conjugated	Sheep	Chemicon
Anti-rabbit biotin	Anti rabbit Ig, biotinylated	Donkey	GE Healthcare

Appendix D *Phytophthora* culture media

Media and solutions used for *Phytophthora* culture in this thesis research are described below. All solutions were stored at RT unless otherwise indicated. All percentages are weight per volume (w/v) unless otherwise stated.

Cleared V8 Juice:

Centrifuged V8 vegetable juice at 10,000 x g for 20 minutes.

Filtered the supernatant through GFA filter paper (Millipore).

Collect cleared V8 juice and stored at –20°C until use.

V8 Agar (1 L)

Cleared V8 juice	100 mL
β-sitosterol	0.02 g
$CaCO_3$	0.1 g
Bacto agar	17 g
ddH_2O	900 mL

Adjusted pH to 6-6.5 with 10 M NaOH. Sterilised by autoclaving at 121°C for 20 min. Poured into shallow 90 mm diameter Petri dishes and then stored in plastic bags at 4°C.

V8 Broth (1 L)

Cleared V8 juice	50 mL
β-sitosterol	0.02 g
$CaCO_3$	0.1 g
ddH_2O	950 mL

Adjusted pH to 6.0-6.5 with 10 M NaOH. Sterilised by autoclaving at 121°C for 20 min.

Appendix E Primers

The oligonucleotide primers used in this thesis research are shown in the table below. All primers were designed using Oligo Analyzer 1.1.0 and Oligo Explorer 1.0.5 primer analysis software (Teemu Kuulasmaa, Finland) and were synthesised at and supplied desalted from Sigma-Genosys, Australia. The oligonucleotides were reconstituted in water following a standard protocol and stored at –20°C as 100 μM stocks.

E.1 Generic Primers

PRIMER NAME	SEQUENCE (5'→ 3')	PURPOSE
T3	AAT TAA CCC TCA CTA AAG GG	pBS vector primer
T7	GTA ATA CGA CTC ACT ATA GGG C	pBS vector primer
Anchor Primer 1	GAC CAC GCG TAT CGA TGT CGA CTT TTT TTT TTT TTT TTA	5' RACE primer
Anchor Primer 2	GAC CAC GCG TAT CGA TGT CGA C	5' RACE primer

E.2 V-PPase Primers

PRIMER NAME	SEQUENCE (5'→ 3')	PURPOSE
DS311 for	TCT TCC TCT GCG GTG TTG TC	**Sequencing** cDNA/genomic clone
DS311 rev	GGA CCA GCG AAC TTT GAG AAC	
PP cDNA rev	GTA GTC CGT GTA GTA TTG TG	
PP cDNA rev2	ACA ACA GTA GAC TTC ATG CCT A	
PP for1	CTG CAT TGA AGC AGA TGC GT	
PP for2	CCA GGC ACT TTC AAT ACT CG	
PP rev1	CTT CAA CAT CAG AGG GAG CTC	
PP rev3	AAC AGT GGC TTC CCT TGG TA	
PP 5'end	TGA TCC CCA CTG CTT TCC GGT	
PP ATG for	ATGCACGAGAAGACACAGG	
PP ATG2 for	GTC GGG TGC GTC TGC GGC	
PP Stop rev	TTATTCACGCGACACGAA	
PP polyT	TTTTTTAACATACTAGTCGAATG	

PP probe 5'end for	GCA CGA GAA GAC ACA GGT TG	**PCR** Probe
PP probe 5'end rev	GTA CTG ACT CCG CCG ACA C	
PP probe 5'end rev2	ACT GAC TCC GCC GAC ACA GA	
PP probe 5'end rev3	GCG GAG CAT ACA GCA CCA AT	
PP probe 5'end rev4	GCT GCC TTA GTG TAA ATA CCG CCA	
PP probe 5'end rev5	CTA CAG AGG ATA CCA GAC CGA GCG	

PRIMER NAME	SEQUENCE (5'→ 3')	PURPOSE
PP 153ins for	ATT CCT CGA CAA CGT GGG CGG	
PP 153ins rev	AAT TCT CGA GGT GCA GCG CAG GCC C	
PP 453ins for	ATA CGG CCC CAT CGC TGA TAA	
PP 453ins rev	AAT TCT CGA GCC CTG GGA AAC GCA	
PP 1244 bp for	AGA TGT TGA AGC TTC CCG TCC CG	**PCR/Cloning**
PP pQE A for	ATC TAT GTG GAT CCG GCC ATG GCA CCA A	Protein
PP pQE A rev	CGA CTG AAC TGC AGG TTA TCC CAC GCG C	overexpression
PP pQE B for	AGG TAT GTG GAT CCG ATG TTG AAG CTT C	
PP pQE B rev	AGA ATG AAC TGC AGG TGC AGC GCA GGC C	
PP pQE C for	ATC TAT GTG GAT CCA AGG AAG TAG CCA CG	
PP pQE C rev	CGA CTG AAG GAT CCG TGC AGC GCA GGC C	
PP dsRNA rev	CCA TCG ATC AAA GAG TCC AGC GT	**PCR/Cloning**

E.3 Aquaporins

PRIMER NAME	SEQUENCE (5'→ 3')	PURPOSE
Aq2_3 for1	TCG CGT AGA TGC ACG CTC TGC	**Sequencing**
Aq2_3 rev1	ATA GCG TAC ACT ACT GGC CC	*Bam*HI 2.3 kb fragment
Aq3_0 for1	TTC CCA AGG CAT TCG TTC CGT	
Aq3_0 for1a	GCG TGT CAG TCA AAG AGC AG	
Aq3_0 for new	ATA CAT GCT GTC GCT ACG	**Sequencing**
Aq3_0 rev1	TGT GTT CAG TGC GTC GAC GT	*Bam*HI 3 kb fragment
Aq3_0 rev2	TGC GGC GAG AGA TAG GCG AA	
Aq3_0 rev3	ATC CAC TTC CAT TCG GTT TGC	
Aq3_0 rev3 new	GTC ACT CCG ATT ATT GAG	
Aq6X for1	GCT GGT GCT TGG CGT GTA	
Aq6X for2	AGC GTT GAT TGA AAA GCA C	
Aq6X for3	AGA GGT CAG CCA CGA GCG-	
Aq6X for4	CCA TAC CTG TTC GAT CAG	**Sequencing**
Aq6X rev1	CCG CAG AAG GCA CCC AAT AG	*Xho*I 6 kb fragment
Aq6X rev2	TGC TGC ACA CAA GTT GAC	
Aq6X rev3	GAC GAA GGT GCC GAG GAA	
Aq6X rev4	GCT AAT GAA CCA ATC GCT	
Aq6X for4 new	AAT GGG CGA GTG GTG GAT	
Aq15S rev1	GTC TTG CGA GTT GGT CAC	
Aq2S for1	GTG CTT GGC GTG TAC GCT A	
Aq2S for2	GAT CGT GGC TCC ACT GTG	**Sequencing**
Aq2S rev1	TAA CAT GGC AGT ACC GAT	*Sal*I 2 kb fragment
Aq2S rev2	CCA TGT TAG TTC TCA GTA	
Aq2S rev3	GTT CTT GTT CGT ATG CTT CG	

PRIMER NAME	SEQUENCE (5'→ 3')	PURPOSE
Aq21S for1	TGG CGG GAG ATG AGT CAG TG	**Sequencing** *SalI* 8 kb fragment
Aq21S for2	GCA GTA CAT GGT GTT GCT	
Aq21S for3	GGT CCT CCT ACT TTT ACT	
Aq21S for4	GGT ATG GCG TTC GGC ATG	
Aq21S for seq-b	GAG AGT ATG GAC ATC GTC	
Aq21S for1 seq-b	TAG GTG GAT GTT GAG CAT	
Aq21S for new1	TAC ATC TAC GAC CAC TGC C	
Aq21S for new2	GTT CCG TCA GTT GTA CCA C	
Aq21S rev1	GAG ATG CTA GAT GAG ACA G	
Aq21S rev2	CTG TCT CAT CTA GCA TCT	
Aq21S rev3	GTG TAG AAG GCA GTA TAG	
Aq21S rev4	TGC CGA GAA ACT CAG CGA	
Aq21S rev5	GAG TAT AAG CAC AGA CCT	
Aq21S rev6	CCT TCT GAT TTG TGT GCT TG	
Aq21S rev7	CTG CTC TTG TAA TGA TGC C	
Aq21S rev seq-a	TAA GCA AAA GTA GAA CTC	
Aq BAC 15S rev	GTA GGC ATA GCG GCG TAG C	**Sequencing**
Aq BAC 21S for	CGA GAG ACG CAG AGA TGG A	
BAC19I9 for1	TAC CGC TTG CTT GTG GAG ATC	
BAC19I9 rev1	GTG GCT ACT ACC GCT CTA TC	

Aq dsRNA rev	CCA TCG ATG AAC TCG GTG TAG	**PCR/Cloning**

Primer Aq1 FOR	GGC TGA CGC CAG ACT CGC	**PCR**
Primer Aq1 REV	AGG TAG CGT CTG GCG CGG	
Primer Aq2 FOR	TCT CAA CAT CCA AAC CGC	
Primer Aq2 REV	CAC TAG CTC TGG CTG AGG C	
Primer Aq3 FOR	TCG ACG CAC TGA ACA CAC	
Primer Aq3 REV	TTG TTG TGG CTG CGG GTG	
Primer Aq4 FOR	GCG TTT GTC CCC ACC GAG	
Primer Aq4 REV	GAG CTG GAG TGG CTG CGG A	
Primer Aq5 FOR	GCG TAC CCT GAT CGA ACA	
Primer Aq5 REV	ATT CTG GTG GGC GAG GGG C	

PS probe 1 for	CTC TTC GGC ATC CAC G	**PCR** *P. sojae* Southern blot probes
PS probe 1 rev	GTG CCC ACC ACC TC	
PS probe 2 for	CAT GGG CGT CTA CTG CT	
PS probe 2 rev	GGT CCG TGA TGG CGT A	
PS probe 3 for	CAT CGT CTA CTA CCC GTG	
PS probe 3 rev	CGG GCT TGT TGA TCT G	
PS probe 4 for	GCC GTG GAA GAA GGT	
PS probe 4 rev	CGG TTG TGC TGG TCC	

SSCP_FOR	AAC TAC ACA GCC TTC TAC AC	**PCR** SSCP analysis
SSCP_REV	ACG AGC GGG ATC CAG AAG T	

Appendix F V-PPase sequence details

Grey areas indicate introns.

F.1 cDNA sequence (DS311, A-5 and E-20)

CCTCTGTAGCTTTCCTCATTGGTGCTGTATGCTCCGCCACAGCAGGATACGTGAGCATGTGGGTCTCGGC
CAGGTCCAATATCCGCGTGGCCAGTGCGGCTCGTGGATCTTATGGAGACGCGCTACTAGTCTGCTTCAGA
GGAGGAGCTTTCTCAGCTGTATTGGACATCACGCTCTGTGTCGGCGGAGTCAGTACACTGTATGTGATGC
TGTACATGCTGTTCGGGTCGGTGCTTCGTCCGACGGAGATCCCGTTGCTGATGGTCGGCTATGGATTCGG
AGCGTCCTTTGTGGCGCTATTCATGCAGCTTGGTGGCGGTATTTACACTAAGGCAGCTGATGTGGGAGCT
GATCTGGTGGGCAAAGTGGAAGTGGGGATTCCAGAGGACGATCCACGTAACCCGGCGGTCATTGCTGACT
TGGTGGGAGATATGGTCGGCGACTGTGTTGGCAGCAGTGCTGATGTGTTCGAGTCTGTTGCTGCAGAGAT
CATTGGAGCGATGATTCTAGGTGGAACTCTGGCGCGTGAGGCTCAACTGCCATATCCTGTGGCTTTTGTA
TTCTTCCCGGTGGTCGTACACGCCTTCGACATTGCTGTGAGCAGCGCTGGGATCTTCATGGTGCGAGCTC
CCTCTGATGTTGAAGCTTCCCGTCCCGATCACAACCCAATGGCTACTCTACAGACCGGCTACAATGTCTC
ACTTTCATTGGCGCTGGTTGGCTTTGCTTTTACCACACGTTGGCTGCTTTACACACCGGAATACCCCTCG
GCATGGATGAACTTCTTCCTCTGCGGTGTTGTCGGCATGCTCACAGCTTACGTGTTTGTCAAGTCGACAC
AATACTACACGGACTACGCTCACCCGCCTGTGCGTAGCATTGCGAAGGCCAGCACGACAGGCCATGGCAC
CAATATCATTACTGGTGTGGCTGTAGGCATGAAGTCTACTGTTGTGCCGACTCTTATGGTGAGTTTCGCT
GTGATATCCGCTTACCACTTGGGTGCCAGCTCAGGTATTGGAGGAGTGGGGAACCGCCACGCTGGACTCT
TTGGAACGGCTGTAGCAACTATGGGAATGCTGTCTTCGGCGGTGTTTGTGCTGGCGATGAACAACTACGG
CCCCATCGCTGATAACGCTGGAGGTATCGCCGAGATGAGTCGTCAGCCGGACTACGTGCGTGATGCTACG
GACAAGCTGGATGCTGCTGGCAATGTGACAAAGGCTATTACTAAGGGATATTCGATCGGCTCGGCTGCTC
TTGCTTGCTTCGTGCTGTTCGGCGCTTTCATGGACGAGTTCTCAGAGTTCGCTGGTCGCGAGTTTAAGAC
GGTGGATATTGCTACGGTGGAGGTCCTGGTTGGCGGTCTGCTGGGTACTATGATGGTCTTCTTCTTTACT
GGACTTGCTGTTGCTGCTGTGGGCGAGACCGCTGGTGAGGTTGTCAATGAAGTGCGCCACCAGTTTGAGA
TTTACCCGGGAATTATGGAGTACAAGGCGAAGCCCGACTACCGCACTTGCGTCGCGCTTGTCACCGAGGC
TGCATTGAAGCAGATGCGTTTCCCAGGGTTGCTTGCTGTGCTGATGCCAGTCTCTGTCGGTATTATCTTC
CGTGTGATTGGTGAGTACCAAGGGAAGCCACTGTTGGGAGCTGAGGCTTTAGCTGGCTACCTCATGTTCG
GTACGGTGACCGGCATCATGATGGCTCTGTTCCTCGACAACGTGGGCGGCGCGTGGGATAACGCCAAGAA
ATACGTCGAGTTGGGCAACTTCGGAGGCAAAGGCAGCGAGGCGCACAAGGCTGCTGTCACTGGCGACACA
GTGGGTGATCCGTTCAAGGATACGGCTGGGCCTGCGCTGCACGTGGTGATCAAGTTGTTGTCCACGACGG
TGCTGGTGTTCGGCCCCCTATTCGTGTCGCGTGAATGAGGGACACTATACAGTTTTGTGGTGCCGAGCTG
GGGACGTAATGACCGTAACCTCGATGAAAAAAAAGGAAATACTGGGTAGCAATGTGCTTAGCTCTCGTGT
CCTACATGTAAACAACGATTTCTAGTCCATTCGACTAGTATGTTAAAAAAAAAAAAAAAAAAACTCGAG

F.2 Genomic clone

GAATTCCTGCAGCCCGGGGGATCCACTTTGGACAGCTCCTCGATGGTCAAGCAGAGACTCATGAATGAGG
CCTCGCTGCCTCCGTATTCCCCTGGGATCTCCACCGACAGAAGTCCATTCTCAAACAGTCCGCGCGTGAT
GGTGTGGTCCATCTCCCCCGCTGTGTCCATGGCACGGACATTAGGCGCCACTACGTCGGCCGCGAATCGG
GCCACAGTGTCCTTGAACATGGTCTCCTCCTCGGAAAGCTCCGTGATGGGGGCGAACGAGGACTCCAGGC
GACGTCCCAGGGGGGAAGGGCGACGCAACGAGCGCGTAAGGCGGTGGGTGGCGGACAACATGGCAAATGA
CTTGAGCTGGACGTGGCAGATTTAAGAGTGACGGTCGCGGGCCCAAAGGTAACTTGAAGTTAATCTACAA
ACTCGTCAAACGATTTTGATAGGATGCACCAATCAGGAACGTTTCTAGGACAGCTGGTTGAAGTTACCTC
GTTATAGTGACAGTTTAGTCATTTCAAGTCAGTCACTATTGCAAGGTAGAAAACAGCAGGTCAGATTCTG
ATCCGCACGCGTGTCAGTCAAAGAGCAGCAATCGAGAGGTCGGACCGTTCACGTCGACGCCAGCTACGTG
ATCCCCACTGCTTTCCGGTGCTGCACTCCGCTCTCTCTGTTCCCAAGGCATTCGTTCCGTGCTCTCCGCT
TCTCTGAACTCCGCCTACAGCCAGCGTCATGCACGAGAAGACACAGGTTGACATCGAGTTCGGCATGTCG
GACACTGCACCGCTGCACAGTGAGCCAGACCTCGAAGAAATGGGGTCGGGTGCGTCTGCGGCCTTGGCTT
ACGCCCGTACACTGGATCTACAGCCTCCGCTCATCGGTATCTCGGTGATAGGAGCGCTGCTGCTGCTGTT
CGGTGGCAGCTTCCGCCCCATCTTTCTGGTCTTCACTATCTGCGGCTACGGCTCCGTGTTCGCGCTCTAC

CTGTCACACTGGGTGCTATCCAAGGACGCAGGCACGACCGAGATGAAGGAAGTAGCCACGCCCATCCGTC
AGGGAGCCGAGGGCTTTCTCAAAATCCAGTACACGGCCATCGCACGCATCGCTGTGGCCATTGCAGGCTT
AATCTTCTTCAGCTACGCACTAAGACCCAGCAGCACGTTGAGCTCTGGTGTGGAGAAGCTCGGCAACTTC
ACGCTCGGTCTGGTATCCTCTGTAGCTTTCCTCATTGGTGCTGTATGCTCCGCCACAGCAGGATACGTGA
GCATGTGGGTCTCGGCCAGGTCCAATATCCGCGTGGCCAGTGCGGCTCGTGGATCTTATGGAGACGCGCT
ACTAGTCTGCTTCAGAGGAGGAGCTTTCTCAGCTGTATTGGACATCACGCTCTGTGTCGGCGGAGTCAGT
ACACTGTATGTGATGCTGTACATGCTGTTCGGGTCGGTGCTTCGTCCGACGGAGATCCCGTTGCTGATGG
TCGGCTATGGATTCGGAGCGTCCTTTGTGGCGCTATTCATGCAGCTTGGTGGCGGTATTTACACTAAGGC
AGCTGATGTGGGAGCTGATCTGGTGGGCAAAGTGGAAGTGGGGATTCCAGAGGACGATCCACGTAACCCG
GCGGTCATTGCTGACTTGGTGGGAGATATGGTCGGCGACTGTGTTGGCAGCAGTGCTGATGTGTTCGAGT
CTGTTGCTGCAGAGATCATTGGAGCGATGATTCTAGGTGGAACTCTGGCGCGTGAGGCTCAACTGCCATA
TCCTGTGGCTTTTGTATTCTTCCCGGTGGTCGTACACGCCTTCGACATTGCTGTGAGCAGCGCTGGGATC
TTCATGGTGCGAGCTCCCTCTGATGTTGAAGCTTCCCGTCCCGATCACAACCCAATGGCTACTCTACAGA
CCGGCTACAATGTCTCACTTTCATTGGCGCTGGTTGGCTTTGCTTTTACCACACGTTGGCTGCTTTACAC
ACCGGAATACCCCTCGGCATGGATGAACTTCTTCCTCTGCGGTGTTGTCGGCATGCTCACAGCTTACGTG
TTTGTCAAGTCGACACAATACTACACGGACTACGCTCACCCGCCTGTGCGTAGCATTGCGAAGGCCAGCA
CGACAGGCCATGGCACCAATATCATTACTGGTGTGGCTGTAGGCATGAAGTCTACTGTTGTGCCGACTCT
TATGGTGAGTTTCGCTGTGATATCCGCTTACCACTTGGGTGCCAGCTCAGGTATTGGAGGAGTGGGGAAC
CGCCACGCTGGACTCTTTGGAACGGCTGTAGCAACTATGGGAATGCTGTCTTCGGCGGTGTTTGTGCTGG
CGATGAACAACTACGGCCCCATCGCTGATAACGCTGGAGGTATCGCCGAGATGAGTCGTCAGCCGGACTA
CGTGCGTGATGCTACGGACAAGCTGGATGCTGCTGGCAATGTGACAAAGGCTATTACTAAGGGATATTCG
ATCGGCTCGGCTGCTCTTGCTTGCTTCGTGCTGTTCGGCGCTTTCATGGACGAGTTCTCAGAGTTCGCTG
GTCGCGAGTTTAAGACGGTGGATATTGCTACGGTGGAGGTCCTGGTTGGCGGTCTGCTGGGTACTATGAT
GGTCTTCTTCTTTACTGGACTTGCTGTTGCTGCTGTGGGCGAGACCGCTGGTGAGGTTGTCAATGAAGTG
CGCCACCAGTTTGAGATTTACCCGGGAATTATGGAGTACAAGGCGAAGCCCGACTACCGCACTTGCGTCG
CGCTTGTCACCGAGGCTGCATTGAAGCAGATGCGTTTCCCAGGGTTGCTTGCTGTGCTGATGCCAGTCTC
TGTCGGTATTATCTTCCGTGTGGTACGTTATTGTACAGAGGGTTTGTTGGTATTCTTGTGTAGTGATGCT
AACCTTGTCTTCCTTTTGGTAGATTGGTGAGTACCAAGGGAAGCCACTGTTGGGAGCTGAGGCTTTAGCT
GGCTACCTCATGTTCGGTACGGTGACCGGCATCATGATGGCTCTGTTCCTCGACAACGTGGGCGGCGCGT
GGGATAACGCCAAGAAATACGTCGAGTTGGGCAACTTCGGAGGCAAAGGCAGCGAGGCGCACAAGGCTGC
TGTCACTGGCGACACAGTGGGTGATCCGTTCAAGGATACGGCTGGGCCTGCGCTGCACGTGGTGATCAAG
TTGTTGTCCACGACGGTGCTGGTGTTCGGCCCCCTATTCGTGTCGCGTGAATAAGGGACACTATACAGTT
TTGTGGTGCCGAGCTGGGGACGTAATGACCGTAACCTCGATGAAAAAAAAGGAAATACTGGGTAGCAATA
TGCTTAGCTCTCGTGTCCTACATGTAAACAACGATTTCTAGTCCATTCGACTAGTATGTTCTCTTGTTAT
CAGGCTCTCGATATTGAGAGTGGATAGACCGGATTGTTAAGTTCTTGAAGGGATTCCTCACGTATAAAGG
CCTCGTCGTCGCAACTGGGTCTGGTGGTACTCGAGTATTCAAGTCTGCCAGAAATCTCCGAGCGACGACT
TGTGCGAGTTGACGCAGGTCTCCATTTCCAGGCACTTTCAATACTCGAACGACGTGGTAAATTCCAGTGA
AAAGCTCTAAGCTGAGGAAAAGAGCCGCTCGGAGATGTTGAGTGCTTGGCCGTGGAGGGTTAGTATTCGA
TGGTGTAAGTGCCATAACGCAGTCATAAATACTGTTTTCTCCATTGTGCTCGCCTTCAACGAGGAAACAC
AGCCATTAAAAAATCA

F.3 *PnVPP* gene with intron

ATGCACGAGAAGACACAGGTTGACATCGAGTTCGGCATGTCGGACACTGCACCGCTGCACAGTGAGCCAG
ACCTCGAAGAAATGGGGTCGGGTGCGTCTGCGGCCTTGGCTTACGCCCGTACACTGGATCTACAGCCTCC
GCTCATCGGTATCTCGGTGATAGGAGCGCTGCTGCTGCTGTTCGGTGGCAGCTTCCGCCCCATCTTTCTG
GTCTTCACTATCTGCGGCTACGGCTCCGTGTTCGCGCTCTACCTGTCACACTGGGTGCTATCCAAGGACG
CAGGCACGACCGAGATGAAGGAAGTAGCCACGCCCATCCGTCAGGGAGCCGAGGGCTTTCTCAAAATCCA
GTACACGGCCATCGCACGCATCGCTGTGGCCATTGCAGGCTTAATCTTCTTCAGCTACGCACTAAGACCC
AGCAGCACGTTGAGCTCTGGTGTGGAGAAGCTCGGCAACTTCACGCTCGGTCTGGTATCCTCTGTAGCTT
TCCTCATTGGTGCTGTATGCTCCGCCACAGCAGGATACGTGAGCATGTGGGTCTCGGCCAGGTCCAATAT
CCGCGTGGCCAGTGCGGCTCGTGGATCTTATGGAGACGCGCTACTAGTCTGCTTCAGAGGAGGAGCTTTC
TCAGCTGTATTGGACATCACGCTCTGTGTCGGCGGAGTCAGTACACTGTATGTGATGCTGTACATGCTGT
TCGGGTCGGTGCTTCGTCCGACGGAGATCCCGTTGCTGATGGTCGGCTATGGATTCGGAGCGTCCTTTGT
GGCGCTATTCATGCAGCTTGGTGGCGGTATTTACACTAAGGCAGCTGATGTGGGAGCTGATCTGGTGGGC
AAAGTGGAAGTGGGGATTCCAGAGGACGATCCACGTAACCCGGCGGTCATTGCTGACTTGGTGGGAGATA
TGGTCGGCGACTGTGTTGGCAGCAGTGCTGATGTGTTCGAGTCTGTTGCTGCAGAGATCATTGGAGCGAT
GATTCTAGGTGGAACTCTGGCGCGTGAGGCTCAACTGCCATATCCTGTGGCTTTTGTATTCTTCCCGGTG
GTCGTACACGCCTTCGACATTGCTGTGAGCAGCGCTGGGATCTTCATGGTGCGAGCTCCCTCTGATGTTG

AAGCTTCCCGTCCCGATCACAACCCAATGGCTACTCTACAGACCGGCTACAATGTCTCACTTTCATTGGC
GCTGGTTGGCTTTGCTTTTACCACACGTTGGCTGCTTTACACACCGGAATACCCCTCGGCATGGATGAAC
TTCTTCCTCTGCGGTGTTGTCGGCATGCTCACAGCTTACGTGTTTGTCAAGTCGACACAATACTACACGG
ACTACGCTCACCCGCCTGTGCGTAGCATTGCGAAGGCCAGCACGACAGGCCATGGCACCAATATCATTAC
TGGTGTGGCTGTAGGCATGAAGTCTACTGTTGTGCCGACTCTTATGGTGAGTTTCGCTGTGATATCCGCT
TACCACTTGGGTGCCAGCTCAGGTATTGGAGGAGTGGGGAACCGCCACGCTGGACTCTTTGGAACGGCTG
TAGCAACTATGGGAATGCTGTCTTCGGCGGTGTTTGTGCTGGCGATGAACAACTACGGCCCCATCGCTGA
TAACGCTGGAGGTATCGCCGAGATGAGTCGTCAGCCGGACTACGTGCGTGATGCTACGGACAAGCTGGAT
GCTGCTGGCAATGTGACAAAGGCTATTACTAAGGGATATTCGATCGGCTCGGCTGCTCTTGCTTGCTTCG
TGCTGTTCGGCGCTTTCATGGACGAGTTCTCAGAGTTCGCTGGTCGCGAGTTTAAGACGGTGGATATTGC
TACGGTGGAGGTCCTGGTTGGCGGTCTGCTGGGTACTATGATGGTCTTCTTCTTTACTGGACTTGCTGTT
GCTGCTGTGGGCGAGACCGCTGGTGAGGTTGTCAATGAAGTGCGCCACCAGTTTGAGATTTACCCGGGAA
TTATGGAGTACAAGGCGAAGCCCGACTACCGCACTTGCGTCGCGCTTGTCACCGAGGCTGCATTGAAGCA
GATGCGTTTCCCAGGGTTGCTTGCTGTGCTGATGCCAGTCTCTGTCGGTATTATCTTCCGTGTGGTACGT
TATTGTACAGAGGGTTTGTTGGTATTCTTGTGTAGTGATGCTAACCTTGTCTTCCTTTTGGTAGATTGGT
GAGTACCAAGGGAAGCCACTGTTGGGAGCTGAGGCTTTAGCTGGCTACCTCATGTTCGGTACGGTGACCG
GCATCATGATGGCTCTGTTCCTCGACAACGTGGGCGGCGCGTGGGATAACGCCAAGAAATACGTCGAGTT
GGGCAACTTCGGAGGCAAAGGCAGCGAGGCGCACAAGGCTGCTGTCACTGGCGACACAGTGGGTGATCCG
TTCAAGGATACGGCTGGGCCTGCGCTGCACGTGGTGATCAAGTTGTTGTCCACGACGGTGCTGGTGTTCG
GCCCCCTATTCGTGTCGCGTGAATAA

F.4 PnVPP protein (801 aa)

MHEKTQVDIEFGMSDTAPLHSEPDLEEMGSGASAALAYARTLDLQPPLIGISVIGALLLLFGGSFRPIFL
VFTICGYGSVFALYLSHWVLSKDAGTTEMKEVATPIRQGAEGFLKIQYTAIARIAVAIAGLIFFSYALRP
SSTLSSGVEKLGNFTLGLVSSVAFLIGAVCSATAGYVSMWVSARSNIRVASAARGSYGDALLVCFRGGAF
SAVLDITLCVGGVSTLYVMLYMLFGSVLRPTEIPLLMVGYGFGASFVALFMQLGGGIYTKAADVGADLVG
KVEVGIPEDDPRNPAVIADLVGDMVGDCVGSSADVFESVAAEIIGAMILGGTLAREAQLPYPVAFVFFPV
VVHAFDIAVSSAGIFMVRAPSDVEASRPDHNPMATLQTGYNVSLSLALVGFAFTTRWLLYTPEYPSAWMN
FFLCGVVGMLTAYVFVKSTQYYTDYAHPPVRSIAKASTTGHGTNIITGVAVGMKSTVVPTLMVSFAVISA
YHLGASSGIGGVGNRHAGLFGTAVATMGMLSSAVFVLAMNNYGPIADNAGGIAEMSRQPDYVRDATDKLD
AAGNVTKAITKGYSIGSAALACFVLFGAFMDEFSEFAGREFKTVDIATVEVLVGGLLGTMMVFFFTGLAV
AAVGETAGEVVNEVRHQFEIYPGIMEYKAKPDYRTCVALVTEAALKQMRFPGLLAVLMPVSVGIIFRVIG
EYQGKPLLGAEALAGYLMFGTVTGIMMALFLDNVGGAWDNAKKYVELGNFGGKGSEAHKAAVTGDTVGDP
FKDTAGPALHVVIKLLSTTVLVFGPLFVSRE

F.5 JGI DB analysis – *P. sojae*

estExt_fgenesh1_pm.C_90025 (PsVPP1)	Protein ID 108429

DNA sequence

```
ATGGCGGAAGCTTTCGAAGCCTACGCCTACGTGGTGAGCGCGGGTGTCTTCGGCATCCTGTTCGCGGC
CTACCTCTTCTGGGAAGTGTCCAAGGTCAAGGTTACGCGCCGCGGTGATGGCTACGCGCTGCTGGCGT
CCGAGGTGCGCCACCAGACGGCCGACCGCCTGTTCGAGATCTTCTGCGCCATCCAGGAGGGCGCCCGC
GCCTTCCTGCTGGCCGAGTACACGCTGTGCTTCGCCTTCATCATCGTCTTCGGCGCCGTCGTGCTCGT
GCTCACGTCGTTCGTCAACAAGGACGGCAAGCAGTTCGACTGGCTCTTCGGCACGCTGAACGCCACGG
CCTTCGCTGTCGGCGGCCTCACGTCCATGGCCGCCGGCTACATGGGCATGATGGTTGCCGTCTACTCG
AACGCCCGTACCACGGTCTCGGCCATGAAGGAGGGCGCCCGCGGCTGGCGCGACGCCTTCAACACGGC
CTTCCGTGCCGGTGCCGTCATGGGCTTCGGACTGAGCTCCATGGCCCTGCTGGTGCTGTTCATCCTGA
TCAAGGCCTTCGAGACGCAGTACCCGCTGTCGACCGACTACAAGAAGCTGTTCGAGGCTATCTCCGGC
TACGGCCTGGGCGGTTCCTCGATTGCCATGTTCGGCCGTGTGGGTGGCGGTATCTACACCAAGGCTGC
CGACGTCGGTGCCGACCTTGCCGGCAAGGTCGTGGAGAACATCCCCGAGGACGACCCGCGTAACCCCG
CTACCATTGCTGACAACGTCGGCGACAACGTCGGTGACGTTGCTGGTATGGGCTCGGACCTTTTCGGC
TCGCTCGCCGAGTCCACGTGCGCGTGCCTGGTGATCTCGACGCAGAGCCCCGAGATCATTGGCGCTGG
CTGGCCCGCTGTCCTGTTCCCTCTGGTGATTACGGCTACGGGCATCTTCGTGTCCGCCATCGTCAGCT
TCCTCGCCACCCACGTGTGGCCCGTCAAGAAGGAGAAGGACGTCGAGACTGTGCTGAAGGTGCAGCTG
TTCGGCTCCACTCTGCTCATGACCATCCTGATCATCCCCGTTGCCCTGTGGCTGCTGCCGTCGACCTT
CTCGATCGGTGGCTCGTTCCAGGTCACCCCGATCCGCGCCTTCTACTGCGTTGCCGTCGGCCTGTGGG
GCGGCTGCATCGTCGGCTTCGTGACGGAGTACTTCACCTCGCACAGCTACCACCCCGTGCGTGAGGTT
GCCCAGGCTTGCGAGACTGGCGCCGCCACCAACATCATCTACGGTCTTGCTCTTGGTTACAAGTCGGC
TATCATCCCGATCACGATCATCTCGCTCGCTGTGTACGTCGGCTTCTCGATGGCCGGCATGTACGGTG
TTGCCCTTGCCGCTCTCGGCTTCCTGGGCACACTCGCTACTTGCCTGGCCATTGATGTGTACGGCCCC
ATTTGTGACAACGCTGGTGGCATTGCCGAGATGGCTGAGCTGCCCGCTGAGGTGCGTGACAAGACCGA
CGCTCTGGACGCTGCCGGTAACACGACGGCTGCTATCGGCAAGGGTTTCGCCATTGGTTCGGCTGCTC
TTGTGTCGCTGGCCCTGTTCGGTGGCTTCGTGACCCGTATTGAGGAGACCTCGATCAACATCCTGTCC
CCGATCACGTTCGCTGGCCTGTTCATGGGTGCTATGCTGCCGTACTGGTTCACCGCCCTGACGATGAA
GTCGGTTGGTGTTGCCGCAATGGAGATGGTCAAGGAGGTGAAGCACCAGTTCGCCACGATCCCCGGTC
TTCTTGAGGGTCTCCCCGGCCACGGTCCTCCGGACCACGCCCGCTGCATCAAGATCTCGACGGACGCC
TCTCTGCGTGAGATGATCCCGCCTGGACTTCTTGTCATGCTGTCCCCCATCGTTGCCGGCACCTTCTT
CGGTGTGCACGCCGTGTCTGGTCTGCTTGTCGGCGGCCTGACCTCGGGTGTGCAGCTCGCCATTTCAC
AGTCGAACACTGGTGGCGCCTGGGACAACGCCAAGAAGTTCGTGGAGAAGGGCTGCGTGTCCATTGAG
GACAAGGACGGCAAGCTCATCGTGCAAGGCAAGGGCAGCGCCATCCACAAGGCCGCCGTTATCGGTGA
CACCGTGGGTGACCCGCTGAAGGATACCTCGGGCCCCGCCCTTAACATTCTGATGAAGCTGATGGCTA
TCATCTCGTTGGTGTTCGGCGACTTCTTCAAGAGCATCAACAACGGCCGCGGTCTGCTGAACGTGCCC
GGCGATGTTGTGTCCGAGGTCGCTGCGGCCCCCGTTGTGGAGGCCACGGCTGCCCCCGCGCTGTAA
```

Protein (804 aa)

```
MKGSSHAGRWFPSPFDPQDPSIFVSAFCRGDLKSTMAEAFEAYAYVVSAGVFGILFAAYLFWEVSKVK
VTRRGDGYALLASEVRHQTADRLFEIFCAIQEGARAFLLAEYTLCFAFIIVFGAVVLVLTSFVNKDGK
QFDWLFGTLNATAFAVGGLTSMAAGYMGMMVAVYSNARTTVSAMKEGARGWRDAFNTAFRAGAVMGFG
LSSMALLVLFILIKAFETQYPLSTDYKKLFEAISGYGLGGSSIAMFGRVGGGIYTKAADVGADLAGKV
VENIPEDDPRNPATIADNVGDNVGDVAGMGSDLFGSLAESTCACLVISTQSPEIIGAGWPAVLFPLVI
TATGIFVSAIVSFLATHVWPVKKEKDVETVLKVQLFGSTLLMTILIIPVALWLLPSTFSIGGSFQVTP
IRAFYCVAVGLWGGCIVGFVTEYFTSHSYHPVREVAQACETGAATNIIYGLALGYKSAIIPITIISLA
VYVGFSMAGMYGVALAALGFLGTLATCLAIDVYGPICDNAGGIAEMAELPAEVRDKTDALDAAGNTTA
AIGKGFAIGSAALVSLALFGGFVTRIEETSINILSPITFAGLFMGAMLPYWFTALTMKSVGVAAMEMV
KEVKHQFATIPGLLEGLPGHGPPDHARCIKISTDASLREMIPPGLLVMLSPIVAGTFFGVHAVSGLLV
GGLTSGVQLAISQSNTGGAWDNAKKFVEKGCVSIEDKDGKLIVQGKGSAIHKAAVIGDTVGDPLKDTS
GPALNILMKLMAIISLVFGDFFKSINNGRGLLNVPGDVVSEVAAAPVVEATAAPAL
```

estExt_Genewise1.C_900043 (PsVPP2)	Protein ID 121300

DNA sequence

```
ATGCTCGAGAAGCCGCCGACCGCCGTGGACATCGAGTTCGGCCTCCCGGACGCCGCGTCCGACCCGCT
GACGGGCGGCGCCGCCTCCGACCCGGACCTGGACGACATGGACATGGGGGCGGGCAGCGCCTCCGCCG
CGCTGGCCTACGCGCGCTCGCTGGACCTGCAGCCGGCGCTCATCGGCATCTCGGCGCTGGGCGCGCTG
CTGCTGCTGTTCGGCGGCAGCTTCCGCCCCATCTTCCTCGTGTTCACCATCTGCGGCTACGGCGCCGC
CTTCGCGCTCTACCTGTCGCACTGGGTGCTGTCCAAGGACGCGGGCTCGGCCGAGATGAAGGAGGTGG
CCACGCCCATCCGCCAGGGCGCCGAGGGCTTCCTGCGCATCCAGTACTCGGCCATCGCGCGCATCGCC
GTGGCCATCGCCGGCCTCATCTTCTTCAGCTACGCGCTGCGCCCCAGCAGCACGCTCGGCTCGGGCGT
CGAGAAGCTCGGCAACTTCACGCTCGGGCTCGTGGCCTCCGCCGCCTTCCTGGTGGGCGCCGTCTGCT
CGGCGGCCGCGGGCTACGTCAGCATGTGGGTCTCGGCCAGGTCCAACATCCGCGTTGCCAGCGCCGCG
CGCGGCAGCTACGGGGAGGCGCTGCTCGTGTGCTTCCGGGGCGGCGCGTTCTCAGCCGTGCTGGACAT
CACGCTCTGCGTGGGCGGCGTCAGCACGCTGTACCTGACGCTGTACATGCTGTTCGGCTCGCTGCTCC
GGCCGACGGAGATCCCGCTGCTCATGGTCGGCTACGGCTTCGGCGCGTCCTTTGTGGCTCTGTTCATG
CAGCTGGGAGGCGGCATTTACACCAAGGCGGCCGATGTGGGGGCAGATCTGGTGGGCAAAGTGGAGGT
GGGCATCCCCGAAGACGACCCGCGCAACCCGGCGGTGATCGCCGACCTGGTGGGAGACATGGTGGGTG
ACTGCGTTGGCAGCAGTGCCGATGTGTTTGAATCCGTCGCGGCCGAGATCATTGGAGCCATGATCCTG
GGCGGCACTCTGGCCCGTGAGGCTCAGCTGCCGTACCCGATCGCCTTTGTGTTCTTCCCTGTGGTTGT
GCACGCGTTCGACATTGCAGTGAGCAGCGCCGGCATTTTGATGGTGAGAGCGCCCTCGGACGTTGAGG
CTTCGCGTCCCGATCACAACCCGATGGCTACGCTGCAGACGGGATACAATGTGTCACTGGCATTGGCG
TTGGTGGGCTTTGCCTTCACCACGCGCTGGCTGCTGTACACGCCCGAGTACCCCACGGCGTGGATGAA
CTTTTTCCTCTGTGGCGTTGTCGGCATGCTCACGGCCTACGTGTTCGTCAAGTCAACGCAGTACTACA
CGGACTACGCGCACCCGCCCGTGCGCAGTATCGCCAAGGCCAGCACCACGGGCCATGGCACGAACATA
ATTACCGGCGTGGCGGTGGGCATGAAGTCGACTGTTGTGCCCACGCTCATGGTGAGCTTCGCCGTGAT
TTCTGCTTATCACTTGGGTGCCGGCTCCGGCATCGGTGGCCAGGGCAACCGGCACGCGGGACTGTTTG
GAACGGCCGTGGCGACTATGGGCATGTTGTCTTCTGCGGTGTTTGTGTTGGCTATGAACAACTACGGG
CCAATCGCGGATAACGCTGGTGGCATTGCTGAGATGAGTCGTCAGCCGGACTATGTGCGAGACGCCAC
GGACAAGTTGGACGCTGCTGGCAATGTTACCAAGGCTATTACGAAGGGTTACTCGATCGGCTCGGCTG
CTCTGGCTTGCTTCGTGCTGTTCGGTGCATTTATGGACGAGTTCTCCGAGTTCGCTGGTCGCGAGTTC
AAGACGGTGGACATTGCGACGGTCGAGGTCCTGGTTGGTGGCCTTTTGGGCACGATGATGGTGTTCTT
CTTCACGGGCCTCGCTGTTGCCGCTGTGGGTGAGACCGCTGGTGAGGTTGTCAATGAAGTGCGCCACC
AGTTCGAGATTTATCCGGGAATCATGGAGTACAAGGCCAAACCCGACTACCGAACGTGCGTTGCGCTG
GTCACCAAGGCAGCATTGAAGCAAATGCGTTTCCCGGGCCTGCTGGCGGTGCTGATGCCAGTATCTGT
TGGCATTGTCTTCCGCGTGGTACGTAACTTTGCTTTAGAGTGACATCGTGTTTGATGCTGACTATTGT
GCACAATTTTGATGTGTAGATTGGAGAGTATCAGGGCAAGCCGCTGCTGGGAGCTGAGGCCCTGGCTG
GCTACCTCATGTTCGGCACGGTTACTGGCATCATGATGGCTCTGTTCCTGGATAACGTGGGCGGAGCC
TGGGACAACGCGAAGAAGTACGTGGAGCTGGGCAACTTCGGAGGCAAGGGAAGCGAAGCGCACAAGGC
CGCGGTCACTGGTGACACGGTGGGTGATCCGTTTAAGGACACTGCTGGTCCGGCCCTGCACGTGGTCA
TCAAGTTGCTGTCCACGACGGTGTTGGTGTTCGGCCCGCTGTTTGTGTCGCGCGAGTAG
```

Protein (812 aa)

```
MLEKPPTAVDIEFGLPDAASDPLTGGAASDPDLDDMDMGAGSASAALAYARSLDLQPALIGISALGAL
LLLFGGSFRPIFLVFTICGYGAAFALYLSHWVLSKDAGSAEMKEVATPIRQGAEGFLRIQYSAIARIA
VAIAGLIFFSYALRPSSTLGSGVEKLGNFTLGLVASAAFLVGAVCSAAAGYVSMWVSARSNIRVASAA
RGSYGEALLVCFRGGAFSAVLDITLCVGGVSTLYLTLYMLFGSLLRPTEIPLLMVGYGFGASFVALFM
QLGGGIYTKAADVGADLVGKVEVGIPEDDPRNPAVIADLVGDMVGDCVGSSADVFESVAAEIIGAMIL
GGTLAREAQLPYPIAFVFFPVVVHAFDIAVSSAGILMVRAPSDVEASRPDHNPMATLQTGYNVSLALA
LVGFAFTTRWLLYTPEYPTAWMNFFLCGVVGMLTAYVFVKSTQYYTDYAHPPVRSIAKASTTGHGTNI
ITGVAVGMKSTVVPTLMVSFAVISAYHLGAGSGIGGQGNRHAGLFGTAVATMGMLSSAVFVLAMNNYG
PIADNAGGIAEMSRQPDYVRDATDKLDAAGNVTKAITKGYSIGSAALACFVLFGAFMDEFSEFAGREF
KTVDIATVEVLVGGLLGTMMVFFFTGLAVAAVGETAGEVVNEVRHQFEIYPGIMEYKAKPDYRTCVAL
VTKAALKQMRFPGLLAVLMPVSVGIVFRVIGEYQGKPLLGAEALAGYLMFGTVTGIMMALFLDNVGGA
WDNAKKYVELGNFGGKGSEAHKAAVTGDTVGDPFKDTAGPALHVVIKLLSTTVLVFGPLFVSRE
```

F.6 JGI DB analysis – *P. ramorum*

fgenesh1_pm.C_scaffold_2000017 (PrVPP1)	Protein ID 71076

DNA sequence

```
ATGGCGGAAGCTTTCGAAGCGTACGCCTACATTGTGAGCGCGGGTGTCTTCGGCATCCTGTTCGCGGC
CTACCTGTTCTGGGAAGTGTCCAAGGTCAAGGTGACGCGTCGCGGCGATGGCTACGCGCTGCTGGCGT
CCGAGGTGCGCCACCAGACGGCCGACCGCCTGTTTGAGATCTACGTGGCCATCCAGGAGGGCGCGCGC
GCCTTCCTGCTGGCTGAGTACACGCTGTGCTTCGCCTTCATCCTCGTCTTCGGCGCTGTGGTGCTCGT
GCTCACGTCGTTCGTGAACAAGGACGGCCAGAAGTTCGACTGGCTCTTCGGCACGCTGAACGCCACGG
CCTTCGCTGTGGGTGGGCTCACGTCCATGGCCGCCGGCTACATGGGCATGATGGTGGCCGTCTACTCG
AACGCGCGTACCACGGTCTCAGCCATGAAGGAGGGCGCTCGCGGCTGGCGCGACGCCTTCAACACGGC
CTTCCGCGCCGGTGCCGTCATGGGTTTCGGACTGAGCTCCATGGCTCTGCTGGTGCTGTTCATCCTGA
TCAAGGCCTTCGAGACGCAGTACCCGCTGTCGACCGACTACAAGAAGCTGTTCGAGGCCATCTCGGGC
TACGGTCTGGGCGGCTCCTCGATTGCCATGTTCGGCCGTGTGGGCGGTGGCATCTACACCAAGGCTGC
CGACGTCGGTGCCGACCTTGCCGGTAAGGTGGTGGAGAACATCCCCGAGGACGACCCGCGTAACCCCG
CTACCATTGCCGACAACGTCGGTGACAACGTTGGTGACGTTGCCGGTATGGGCTCGGACTTGTTCGGC
TCGCTTGCCGAGGCTACGTGTGCCTGCCTCGTTATCTCGACGCAGAGCCCCGAGATCATCGGCGCTGG
CTGGCCCGCCGTTCTGTTCCCGCTCGTGATCACGGCTACGGGTATCTTCGTGTCCGCCATCATTAGCT
TCCTCGCCACCCACGTGTGGCCCGTCAAGAAGGAGAAGGACGTCGAGACTGTGCTGAAGGTGCAGATC
TTCGGCTCCACTCTGCTCATGACTATTCTCATCGTTCCGGTTGCCCTGTGGCTGCTGCCGTCGACTTT
CTCGATCGGCGGTACGTACCAGGTGACCCCTATCCGTGCCTTTTACTGCGTTGCCGTCGGTCTTTGGG
GTGGCTGCATCGTCGGCTTCGTGACGGAGTACTTCACGTCGCACAGCTACCAGCCCGTGCGCGAGGTT
GCCCAGGCTTGCGAGACTGGTGCTGCTACGAACATCATCTATGGTCTTGCTCTTGGTTACAAGTCGGC
TATTATCCCGATCACGATCATCTCGATCGCTGTGTACATCGGCTTCTCGACGGCTGGCATGTACGGTG
TTGCTCTTGCTGCTCTTGGCTTCCTGGGCACGCTCGCTACCTGCCTGGCCATTGATGTCTATGGCCCC
ATCTGTGATAATGCTGGTGGTATTGCCGAGATGGCTGAGTTGCCCGCTGAGGTGCGTGACAAGACTGA
TGCTCTTGACGCTGCCGGTAACACGACGGCTGCTATCGGCAAGGGATTCGCGATTGGCTCGGCCGCTC
TTGTGTCGCTCGCACTGTTCGGAGGTTTCGTGACCCGTATTGAGGAGACGTCTATCAACATCTTGTCG
CCGATCACGTTCGCTGGCCTATTCATGGGTGCCATGCTGCCGTACTGGTTCACCGCTTTGACTATGAA
GTCTGTGGGTGTTGCCGCTATGGAGATGGTCAAGGAAGTCAAGCACCAGTTCGCCACGATCCCCGGTC
TTCTTGAGGGTCTGCCCGGTCACGGTCCTCCGGACCACGCTCGTTGCATCAAGATCTCCACGGACGCT
TCTCTGCGTGAGATGATCCCGCCTGGACTTCTGGTCATGCTGTCCCCCATCATTGCCGGCACCGTGTT
CGGTGTCCACGCCGTGTCTGGCCTGCTTGTCGGCGGCCTGACGTCCGGTGTGCAGCTCGCCATTTCAC
AGTCGAACACTGGTGGCGCCTGGGATAACGCGAAGAAGTTCGTCGAGAAGGGCTGCGTGTCTATTGAG
GACAAGGATGGCAAGCTTATCGTGCAGGGCAAGGGCAGCGCCATCCACAAGGCCGCCGTTGTCGGCGA
CACCGTGGGTGACCCGTTGAAGGATACCTCTGGTCCTGCCCTGAACATTTTGATGAAGCTGATGGCTA
TCATCTCTCTGGTGTTCGGCGACTTCTTCAAGAGCATCAACAACGGCCGTGGCTGCTGCCGCTCCTGC
CGTGGACGTCACGGCCGCTCCCCAGCTGTAA
```

Protein (769 aa)

```
MAEAFEAYAYIVSAGVFGILFAAYLFWEVSKVKVTRRGDGYALLASEVRHQTADRLFEIYVAIQEGAR
AFLLAEYTLCFAFILVFGAVVLVLTSFVNKDGQKFDWLFGTLNATAFAVGGLTSMAAGYMGMMVAVYS
NARTTVSAMKEGARGWRDAFNTAFRAGAVMGFGLSSMALLVLFILIKAFETQYPLSTDYKKLFEAISG
YGLGGSSIAMFGRVGGGIYTKAADVGADLAGKVVENIPEDDPRNPATIADNVGDNVGDVAGMGSDLFG
SLAEATCACLVISTQSPEIIGAGWPAVLFPLVITATGIFVSAIISFLATHVWPVKKEKDVETVLKVQI
FGSTLLMTILIVPVALWLLPSTFSIGGTYQVTPIRAFYCVAVGLWGGCIVGFVTEYFTSHSYQPVREV
AQACETGAATNIIYGLALGYKSAIIPITIISIAVYIGFSTAGMYGVALAALGFLGTLATCLAIDVYGP
ICDNAGGIAEMAELPAEVRDKTDALDAAGNTTAAIGKGFAIGSAALVSLALFGGFVTRIEETSINILS
PITFAGLFMGAMLPYWFTALTMKSVGVAAMEMVKEVKHQFATIPGLLEGLPGHGPPDHARCIKISTDA
SLREMIPPGLLVMLSPIIAGTVFGVHAVSGLLVGGLTSGVQLAISQSNTGGAWDNAKKFVEKGCVSIE
DKDGKLIVQGKGSAIHKAAVVGDTVGDPLKDTSGPALNILMKLMAIISLVFGDFFKSINNGRGLLNVP
LDVLPEAAAAPAVDVTAAPQL
```

fgenesh1_pm.C_scaffold_10000002 (PrVPP2)	Protein ID 71249

DNA sequence

```
ATGCTCGACAAGACGCAGGTGGATATCGAATTCGGCATCGCGGACGATGCTCCGCTCAACTCGGAGTC
CGACCACGATGAGATGGGGGCGGGCGCTTCCGCGGCGCTCGCCTACGCGCGCAGTCTGGACCTGCAGC
CGCCGCTGGTCGCCATCTCGGTCGTGGGGGCGCTGCTGCTTCTCTTCGGCGGCAGCTTCCGCCCCATC
TTTCTCGTGTTCACGCTTTGCGGCTACGGGGCCGTGTTCGCGCTCTTCCTGTCACACTGGGTGCTGGC
CAAAGACGCGGGCTCTGCAGACATGAGGGAGGTGGCCACTCCCATCCGCCAGGGAGCCAAGGGCTTCC
TGCGCATCCAGTACTCGGCCATCGCCCGCATTGCGCTGGCCATCGCCGCCCTCATCTTCTGCAGCTAC
GCGCTGCGCCCCAGCAGCGCCTTAAGCTCTGGCGTGGAGAAACTGGGCAACTTTACGCTGGGCCTCGT
GGCCTCCGTCGCCTTTTTGGTCGGCGCCGTGTGCTCAGCGGCCGCCGGGTACGTGAGTATGTGGGTCT
CAGCCAGGTCCAACGTCCGTGTTACAAGTGCTGCCCGTCGGTCCTACGGAGAGGCGCTGCTGGTGTGC
TTCAGGGGTGGGGCCTTCTCAGCTGTGCTGGACATCACGCTGTGTGTGGGCGGCGTGAGCTCGCTCTA
CGTGACGCTGTACATGATGTTCGGGTCGATCTTGCGCCCTACAGAGATCCCTTTGCTGATGGTGGGGT
ACGGGTTTGGCGCGTCGTTCGTGGCGCTGTTCATGCAGCTTGGAGGGGGGATTTACACCAAGGCGGCC
GATGTGGGCGCCGATTTGGTGGGGAAGGTGGAAGTGGGTATTCCAGAGGATGATCCGCGCAACCCGGC
GGTGATTGCCGACCTCGTGGGAGACATGGTGGGCGACTGCGTGGGCAGCAGTGCCGATGTGTTCGAGT
CTGTGGCTGCGGAGATTATCGGCGCGATGATCTTGGGCGGCACGCTCGCGCGGGAAGCGCAGCTGCCG
TACCCGGTGTCTTTTGTGTTCTTCCCTGTGGTCGTGCACGCCTTTGACATTGCAGTGAGCAGTGCTGG
CATTTTGATGGTGAGAGCACCCTCGGACGTCGAAGCTTTGCGTCCCGATCACAACCCAATGGCCACGA
TGCAGACTGGTTATAATGTATCTCTCGCTCTCGCGCTGGTTGGCTTTGGCGTGACGACGCGCTGGCTG
CTCTACACGCCGGAGCACCCTACAGCGTGGATAAATTTCTTCCTCTGCGGCGTGGTCGGTATGCTCAC
AGCTTACGTGTTTGTTAAGTCGACGCAGTACTACACGGACTACGCGCACCCGCCTGTGCGTAGCATCG
CGAAGGCCAGCACCACGGGCCATGGCACCAATATCATTACCGGCGTGGCTGTGGGCATGAAGTCCACG
GTTGTTCCTACACTTATGGTGAGCTTTGCCGTCATCTCCGCTTACCACTTGGGGGCAGGCTCGGGCAT
TGGCGGCCAGGGGAACCGCCACGCTGGGTTATTCGGCACGGCAGTGGCGACCATGGGCATGCTGTCCT
CGGCAGTGTTTGTGCTTGCTATGAACAACTACGGCCCCATCGCTGATAATGCTGGCGGTATTGCCGAG
ATGAGCCACCAACCGGACTACGTGCGCGACGCTACAGACAAGTTGGACGCTGCTGGGAATGTGACGAA
GGCCATCACCAAGGGCTACTCGATCGGTTCGGCTGCTCTGGCCTGCTTCGTGCTCTTTGGTGCTTTCA
TGGACGAGTTTTCGGAGTTCGCTGGTCGTGAGTTCAAGACGGTGGATATCGCAACGGTCGAGGTCCTC
GTGGGCGGTCTGCTGGGTACCATGATGGTGTTTTTCTTCACGGGGCTTGCTGTTGCCGCCGTGGGTGA
GACCGCTGGCGAGGTTGTGAACGAGGTGCGCCGTCAGTTCGAGATGTACCCGGGCATCATGGAGTACA
AAGCAAAGCCCGACTATCGCACGTGCGTAGCTCTCGTCACCAAGGCTGCCTTGAAGCAGATGCGCTTC
CCAGGTCTGCTGGCTGTGTTGATGCCGGTTTGTGTTGGTATTGTCTTCCGTGTGGTACGTTGTTGTTG
CATATCTTTCCTATAATGTGCACACATGCTGACTGCTGTTTGTGGTGGGTGATATTAGATCGGAGAGT
ACCAGGGCAAGCCACTGCTAGGCGCGGAAGTTCTAGCTGGTTACCTCATGTTCGGCACGGTGACTGGC
ATTATGATGGCTCTGTTCCTGGACAACGTCGGCGGTGCCTGGGACAACGCCAAGAAGGTATGACTGAG
AAATCAGTATTGATGGCTTGCCTTGGGACTAACTCTATTAATTAACGGCATCAAACCGTGGCTCACCC
ATGTTATGTGTCTTGTATGACGACAGTACGTGGAGCTGGGCAACTTTGGAGGCAAAGGGAGTGAGGCG
CACAAGGCTGCTGTGACTGGCGACACGGTGGGCGATCCGTTCAAGGACACGGCCGGCCCAGCGCTGCA
CGTGGTCATCAAGTTGCTGTCCACGACGGTACTGGTGTTTGGCCCGCTGTTTGTGTCGCGCGAGTAA
```

Protein (801 aa)

```
MLDKTQVDIEFGIADDAPLNSESDHDEMGAGASAALAYARSLDLQPPLVAISVVGALLLLFGGSFRPI
FLVFTLCGYGAVFALFLSHWVLAKDAGSADMREVATPIRQGAKGFLRIQYSAIARIALAIAALIFCSY
ALRPSSALSSGVEKLGNFTLGLVASVAFLVGAVCSAAAGYVSMWVSARSNVRVTSAARRSYGEALLVC
FRGGAFSAVLDITLCVGGVSSLYVTLYMMFGSILRPTEIPLLMVGYGFGASFVALFMQLGGGIYTKAA
DVGADLVGKVEVGIPEDDPRNPAVIADLVGDMVGDCVGSSADVFESVAAEIIGAMILGGTLAREAQLP
YPVSFVFFPVVVHAFDIAVSSAGILMVRAPSDVEALRPDHNPMATMQTGYNVSLALALVGFGVTTRWL
LYTPEHPTAWINFFLCGVVGMLTAYVFVKSTQYYTDYAHPPVRSIAKASTTGHGTNIITGVAVGMKST
VVPTLMVSFAVISAYHLGAGSGIGGQGNRHAGLFGTAVATMGMLSSAVFVLAMNNYGPIADNAGGIAE
MSHQPDYVRDATDKLDAAGNVTKAITKGYSIGSAALACFVLFGAFMDEFSEFAGREFKTVDIATVEVL
VGGLLGTMMVFFFTGLAVAAVGETAGEVVNEVRRQFEMYPGIMEYKAKPDYRTCVALVTKAALKQMRF
PGLLAVLMPVCVGIVFRVIGEYQGKPLLGAEVLAGYLMFGTVTGIMMALFLDNVGGAWDNAKKYVELG
NFGGKGSEAHKAAVTGDTVGDPFKDTAGPALHVVIKLLSTTVLVFGPLFVSRE
```

F.7 VMD DB analysis – *P. infestans*

CL1Contig3651 (PiVPP1)

DNA sequence

```
ATGGCGGAAGCTTTTGAAGTCTACGCCTACGTGGTGACCGCGGGTGTGTTCGGCATCCTTTTCGCGGC
CTACCTGTTCTGGGAAGTTTCCAAGGTCAAGGTGACTCGTCGTGGCGATGGCTACGCGCTGCTGGCGT
CGGAGGTGCGCCACCAGACTGCCGACCGTCTGTTCGAGATCTTCTGCGCCATTCAAGAGGGCGCCCGC
GCCTTCTTGCTGGCAGAGTACACGCTGTGCTTTGCCTTCATCATCGTCTTCGGTGCCGTCGTGCTTGT
ACTGACGTCGTTCGTCAACAAGGACGGCAAGCAGTTCGACTGGCTTTTCGGTACGCTGAACGCCACGG
CCTTCGCCGTCGGTGGTCTCACCTCGATGGCCGCGGGTTACATGGGTATGATGGTGGCTGTCTACTCG
AACGCTCGTACCACGGTCTCGGCCATGAAGGAGGGTGCTCGTGGCTGGCGTGACTCGTTCAACACGGC
CTTCCGTGCCGGTGCTGTCATGGGCTTCGGACTGAGCTCCATGGCTCTGCTGGTGCTGTTCATCCTGA
TCAAGGCTTTCGAGACTCAGTACCCTCTGTCGACTGACCACAAGAAGTTGTTCGAGGCCATCTCGGGC
TACGGTCTGGGTGGTTCGTCCATTGCCATGTTCGGCCGTGTGGGTGGCGGTATCTACACCAAGGCTGC
TGATGTCGGTGCTGATCTGGCTGGTAAGGTCGTGGAGAACATCCCTGAGGATGACCCGCGTAACCCCG
CTACCATTGCTGACAATGTCGGTGACAACGTCGGTGACGTCGCCGGTATGGGTTCGGACCTTTTCGGC
TCGCTCGCTGAGGCGACGTGCGCGTGTCTGGTGATCTCGACTCAGAGCCCTGAGATCATTGGCGCTGG
CTGGCCTGCTGTTCTGTTCCCGCTGGTGATCACCGCTACCGGTATTTTCGTGTCTGCAATCATCAGCT
TCCTTGCTACTCACGTGTGGACTGTAAAGAAGGAGAAGGACGTCGAGCTTGTGCTGAAGGTTCAGCTG
TTCGGCTCCACCCTGCTCATGACTGTTTTGATCGTTCCCGTGGCACTGTGGCTGCTGCCATCGCACTT
CTCTATCGGTACGGCGTACCAGGTGACCCCTATCCGTGCTTTCTACTGTGTTGCTGTCGGCCTGTGGG
GCGGCTGCATTGTCGGCTTCGTGACGGAATACTTCACCTCGCACAGCTACAAGCCCGTGCGTGAGGTT
GCTCAGGCTTGCGAGACTGGTGCCGCTACCAACATCATCTACGGTCTTGCTCTGGGCTACAAGTCGGC
TATCATCCCTATCACGATCATTTCCCTGGCTGTATACGTTGGATTCTCGACCGCTGGCATGTACGGTG
TCGCCCTTGCCGCTCTTGGTTTCCTGGGTACGCTAGCCACTTGCCTGGCTATTGATGTGTATGGTCCC
ATCTGTGACAACGCTGGTGGCATTGCCGAGATGGCTGAGCTGCCCGCTGAAGTGCGTGACAAGACTGA
CGCTCTTGACGCTGCTGGTAATACCACGGCTGCTATTGGCAAGGGTTTCGCCATTGGCTCGGCTGCTC
TTGTGTCCCTCGCTCTGTTCGGTGGCTTTGTCACTCGTATCGAAGAGGAATCTATTAACATTTTGAGC
CCGATCACGTTCGCTGGTCTGTTCATGGGTGCCATGCTGCCGTACTGGTTTACTGCCATGACGATGAA
GTCGGTTGGCGTTGCTGCTATGGAGATGGTCAAGGAGGTCAAGCACCAGTTCGCCACGATCCCTGGTC
TTCTTGAGGGTCTGCCTGGCCACGGCCCTCCGGACCACGCCCGTTGCATCAAGATCTCTACCGATGCT
TCTCTGCGTGAGATGATCCCGCCCGGACTTCTTGTGATGCTCTCCCCTATTATTGCCGGCACCTTCTT
CGGTGTCCACGCCGTGTCTGGTCTGCTTGTCGGCAGTCTGACGTCGGGTGTGCAGCTCGCTATCTCGC
AGTCGAACACTGGCGGTGCCTGGGACAACGCCAAGAAGTTCGTGGAGAAGGGATGCGTCTCCATTGAG
GACAAGGAGGGCAAGCTTATTGTCCAGGGCAAGGGAAGTGCCATCCACAAGGCCGCTGTCATTGGTGA
CACCGTGGGTGACCCGTTGAAGGACACTTCTGGTCCGGCCCTTAACATCCTGATGAAGCTTATGGCTA
TCATCTCGCTGGTGTTCGGAGACTTTTTCAAGAGCATCAACAACGGCCGCGGTCTGCTGAACGTGCCC
GCCGATGTGTCGGCTGCTGCTGCCCCCGTTGTGGAGGTCACGGCGGCTCCTCAGCTGTAA
```

Protein (767 aa)

```
MAEAFEVYAYVVTAGVFGILFAAYLFWEVSKVKVTRRGDGYALLASEVRHQTADRLFEIFCAIQEGAR
AFLLAEYTLCFAFIIVFGAVVLVLTSFVNKDGKQFDWLFGTLNATAFAVGGLTSMAAGYMGMMVAVYS
NARTTVSAMKEGARGWRDSFNTAFRAGAVMGFGLSSMALLVLFILIKAFETQYPLSTDHKKLFEAISG
YGLGGSSIAMFGRVGGGIYTKAADVGADLAGKVVENIPEDDPRNPATIADNVGDNVGDVAGMGSDLFG
SLAEATCACLVISTQSPEIIGAGWPAVLFPLVITATGIFVSAIISFLATHVWTVKKEKDVELVLKVQL
FGSTLLMTVLIVPVALWLLPSHFSIGTAYQVTPIRAFYCVAVGLWGGCIVGFVTEYFTSHSYKPVREV
AQACETGAATNIIYGLALGYKSAIIPITIISLAVYVGFSTAGMYGVALAALGFLGTLATCLAIDVYGP
ICDNAGGIAEMAELPAEVRDKTDALDAAGNTTAAIGKGFAIGSAALVSLALFGGFVTRIEEESINILS
PITFAGLFMGAMLPYWFTAMTMKSVGVAAMEMVKEVKHQFATIPGLLEGLPGHGPPDHARCIKISTDA
SLREMIPPGLLVMLSPIIAGTFFGVHAVSGLLVGSLTSGVQLAISQSNTGGAWDNAKKFVEKGCVSIE
DKEGKLIVQGKGSAIHKAAVIGDTVGDPLKDTSGPALNILMKLMAIISLVFGDFFKSINNGRGLLNVP
ADVSAAAAPVVEVTAAPQL
```

CL62Contig1 (PiVPP2a)
DNA sequence
ATCGCCCGCATCGCTGTAGCCATCGCAGGCTTGATCTTCTTCAGTTACGCACTCAGACCCAGCAGCAC GTTGGGCAGCTCGGGCGTGGAGAAGCTCGGCAACTTCACTCTCGGTCTGGTATCTTCTGTGGCTTTCC TCATTGGCGCCGTATGCTCCGCCGCGGCAGGATACGTGAGCATGTGGGTTTCGGCCAGGTCCAACGTC CGTGTGGCTAGTGCCGCTCGAGGATCTTATGGAGAAGCGCTGCTGGTCTGCTTCAGGGGTGGAGCTTT CTCGGCTGTGCTGGATATCACGCTCTGTGTGGGCGGAGTCAGCACGCTGTATGTGATGCTGTACATTC TATTCGGATCTGTGCTGCGCCCCACGGAGATCCCGTTGCTGATGGTCGGCTATGGATTCGGAGCGTCC TTTGTGGCTCTGTTCATGCAGCTGGGTGGCGGTATCTACACTAAAGCAGCAGATGTGGGAGCTGATCT GGTGGGCAAAGTGGAAGTGGGGATCCCAGAGGACGATCCACGTAACCCTGCGGTCATAGCAGACTTGG TGGGGGATATGGTTGGTGACTGTGTTGGCAGCAGTGCTGATGTGTTCGAGTCAGTGGCTGCAGAGATC ATTGGAGCGATGATTTTGGGTGGAACTTTGGCGCGCGAGGCTCAACTGCCGTATCCGGTGGCCTTTGT GTTCTTCCCGGTGGTCGTACACGCCTTTGATATTGCGGTGAGCAGTGCTGGGATCTTCATGGTGCGAG CTCCTTCCGACGTCGAAGCTTGCATGCCTGCAGG
Protein (260 aa)
IARIAVAIAGLIFFSYALRPSSTLGSSGVEKLGNFTLGLVSSVAFLIGAVCSAAAGYVSMWVSARSNV RVASAARGSYGEALLVCFRGGAFSAVLDITLCVGGVSTLYVMLYILFGSVLRPTEIPLLMVGYGFGAS FVALFMQLGGGIYTKAADVGADLVGKVEVGIPEDDPRNPAVIADLVGDMVGDCVGSSADVFESVAAEI IGAMILGGTLAREAQLPYPVAFVFFPVVVHAFDIAVSSAGIFMVRAPSDVEACMPA

CL790Contig1 (PiVPP2b)
DNA sequence
ATGAACTTTTTCCTCTGCGGTGTAGTGGGGATGCTCACAGCCTACGTGTTTGTTAAGTCGACGCAGTA CTACACAGACTACGCTCATCCACCTGTGCGCGGTATTGCGAAGGCCAGCACTACGGGTCACGGAACCA ATATCATTGCCGGCGTGGCTGTGGGCATGAAATCTACCGTTGTACCGACGCTGATGGTGAGTTTCGCC GTGATCTCGGCCTACCACTTGGGTGCCAGATCTGGCATTGGGGGAGTGGGGAACCGCCACGCTGGACT CTTTGGAACGGCCGTGGCGACTATGGGGATGCTATCCTCTGCAGTGTTTGTGCTGGCGATGAACAACT ACGGCCCAATCGCAGACAATGCTGGAGGTATCGCCGAGATGAGTCGTCAACCAGACTACGTGCGTGAT GCTACGGACAAGCTGGATGCTGCAGGCAACGTGACCAAGGCTATTACTAAAGGCTACTCGATCGGCTC GGCTGCTCTGGCTTGCTTCGTGCTGTTCGGCGCTTTCATGGACGAGTTCTCCGAGTTCGCTGGTCGCG AGTTCAAGACGGTGGATATTGCAACGGTGGAGGTTCTGGTCGGTGGTCTGCTGGGTACAATGATGGTC TTCTTCTTCACGGGGCTTGCCGTTGCTGGAGTGGGCGAGACGGCCGGTGAGGTTGTCAACGAAGTGCG CCACCAGTTCAAGATCTACCCGGGTATTATGGAGTACAAGGAGAAACCCGACTACCGCACGTGTGTTG CGCTAGTCACAAAGGCTGCGTTGAAGCAGATGCGCTTCCCAGGTCTGTTGGCTGTGTTGATGCCAGTC TCTGTCGGTGTTGTCTTCCGTGTGGTACGTTTTGCTGTCATGAATGTCTGTTGGTAG
Protein (290 aa)
MNFFLCGVVGMLTAYVFVKSTQYYTDYAHPPVRGIAKASTTGHGTNIIAGVAVGMKSTVVPTLMVSFA VISAYHLGARSGIGGVGNRHAGLFGTAVATMGMLSSAVFVLAMNNYGPIADNAGGIAEMSRQPDYVRD ATDKLDAAGNVTKAITKGYSIGSAALACFVLFGAFMDEFSEFAGREFKTVDIATVEVLVGGLLGTMMV FFFTGLAVAGVGETAGEVVNEVRHQFKIYPGIMEYKEKPDYRTCVALVTKAALKQMRFPGLLAVLMPV SVGVVFRVVRFAVMNVCW

F.8 JGI DB analysis – *T. pseudonana*

newV2.0.genewise.13.39.1 (TpVPP1)	Protein ID 119500

DNA sequence

CACGTCGGCGTCGGGCTTGGTGCACCACATCATCGTTTCATTTCATTGAANTTCAATCATATTNTCCA
CTCAACCACTGATCACCTCCATCCAACCCGATCCCTTCTCTCTTCAAAGGTGTCAGCAACAGCCATCA
TCATCTCAGCCGCCCTCGTGGGCATCGCCTGGGCCCTCCTCCAATTCCTCCTAATCACCCGCATCCCC
GTCCGCTCCGAAGGCATCTCGGACTCCACCGGCCTCGTTAGCGGATCCAACGACGAGGCCACCACCCG
TCGTCTCAATGAGATCTACACCGCCATTTATGAGGGGGCAGAATCCTTCCTTCGTGCCGAGTACATGA
TGTGCGCTTGGTTCGTTTTGGTCTTTTCGGTCATCATCTTTGTGCTCGTTTCTTGGGGAACTGGGTGG
GACTTTGCACGTGGTACTTTCACGGCCTTGTCTTTTATTCTTGGCGCTTGTACTTCCATCCGTAAGTG
TTGTTTTGGGGGGAGGGAGGAGAGATTTTGGGCTTGTTCGCATCTTGTGCCAAACATCTATACTACCG
TGTCGTGTGGCTGACAATACACTCTTTCTTCTCCACTCCTCAGTCTCGGGCTACCTTGGAATGAAGGT
CGCTGTGTACTCCAACGTGAGGACTACCGTATCAGCTCAAAAGTCTGGGTGGAAGCACTGCTTCAACG
CTGCCTTCCGTGCCGGAGCCGTCATGGGATTTGCCCTCTGCGGTCTTGGAATCTTCATGCTTTACGTC
ACTCTCTTGGCATTCCGTGTTCACTACCCTGCTCCTGAGGACTGGATCTACCTCACTGAGTGCTTGAC
TGGATACGGACTTGGAGGAAGTAGTATTGCCATGTTTGGTCGTGTTGGTGGAGGAATCTACACCAAAG
CAGCTGATGTGGGAGCTGATCTTGTTGGAAAGGTCGTTCATGGAATTCCTGAGGATGATCCTCGTAAC
CCGGCTACCATCGTGAGTATCGTTTGTGTTGTTGGGTAGAGTCATTGTATTGAGATGGAAATCTATGC
TCACGATGTGGAAATCGTCCATTACAGGCTGACAACGTGGGAGACAATGTTGGAGATGTTGCCGGCAT
GGGAAGTGATCTTTTCGGTTCCTTCGCTGAGTCTACCTGCGCTGCACTTGTGCTCGGAAGCTCCATCG
GTTTGTCTGGGGGTTGGGACGCTATGGTGTTTCCTGTCGCCGTCTCAGCTGTAGGAATCTTCGTATGC
CTCATCTGCTCCTTCATTGCTACGGACCTTCGCCCTGTGAAGGAAGAGAAGGATGTTGAAATGTCACT
CAAGGTGCAACTCATCAGTACTACGGTTCTCATGATTCCAGCTGTATACCTTGCTGCCGAAACCTTCC
TTCCCGGAGAATTCATGCTTAAGGCTACTGTTGGTAACAACGTGCTTACTCTTCACCCTTGGCAAGCA
TTCCTGTGCGTGATCATGGGAGCCGTTGGAGGTCTCATCATCGGTCTCGTTACTGAGTACTACACTTC
TCACTCCTACAAACCCGTTCGTGAATTGGCAGACTCGTGTAAGACTGGAGCAGCAACCAACATGATCT
ACGGAATTGCTCTTGGATACAAATCAGCTATCATCCCCGTATTGATCTTGGCCGTCGTGGTGTATGGA
TCCTTCTCTATGTGCGACATGTACGGAGTGTCTCTTGCTGCCATCGGATTCCTTTCCAACCTTGCAAC
TGGGCTCACGATTGACGTGTATGGTCCCGTTTGCGACAATGCTGGTGGTATTGCTGAAATGGCGGAAC
TCGAGCCTTATGTTCGAGAAAAGACCGATGCCCTCGATGCCGCGGGAAATACTACCGCCGCTATTGGC
AAGGGATTCGCAATTGGATCGGCTGCTCTTGTATCCCTCGCTCTTTTCGGAGCCTTTGTCACTCGGTA
AGCATTCTTGACTGCATCTTGTCTGTTGAACGTTGCATCTATCCCCTCTCACCCACCCCCTCGCCTCC
GAATAGTATCCGCCATGCCTCCAATGATGAGCTCTTCCAGGACGGTGTTAATATGTTGAGTCCACTCA
CATTTGCCTTCTTGATCATCGGTGGCATGATTCCTTTCGCATTTGCCGCCATGACGATGAAGTCTGTC
GGAGTTGCTGCCATGGAGATGGTGTTAGAGGTTCAGCGTCAGTTCGATGAGAAGCCCCACCTTTTGGA
TGCAAACCCAACTGAACGCCCTGACTATGACGCTTGCATTGGTATCAGTACTAAGGCGTCTCTCAAAG
AAATGGTTGCACCTGGTGCTATGGTTATCTTAACCCCACTCTTGACCGGTATCTTCTTCGGAGTGTAT
GCCGTGTCTGGTCTTCTTGTTGGATCTCTCGTTGCTGCCGTTCAGCTCGCCATTTCCATGTCTAACTC
GGGCGGTGCATGGGATAATGCCAAGAAGTACAGTGAGTGGAAAGATTGTATTTTTTATTGTATTGAAC
TGCGGTTTACATGCAAACTAACTTGTGATTGTGACCATTGTAGTTGAAAAGGCAACCCCTGATTCCGA
TTTGAAGGGCAAAGGAAGTGATATCCACAAAGCTGCAGTTGTTGGAGATACTGTCGGAGATCCTTTTA
AGGATACATCTGGTCCTGCGCTCAACATTGTCATGAAGTTGATGGCAGTGCTTTCTTTGGTGTTTGCT
GACACCTTTTACGCAACGAACGGCGGTGCTGGGGTGTTCCAGCTTTAA

Protein (769 aa)

HVGVGLGAPHHRFISLXFNHIXHSTTDHLHPTRSLLSSKVSATAIIISAALVGIAWALLQFLLITRIP
VRSEGISDSTGLVSGSNDEATTRRLNEIYTAIYEGAESFLRAEYMMCAWFVLVFSVIIFVLVSWGTGW
DFARGTFTALSFILGACTSILSGYLGMKVAVYSNVRTTVSAQKSGWKHCFNAAFRAGAVMGFALCGLG
IFMLYVTLLAFRVHYPAPEDWIYLTECLTGYGLGGSSIAMFGRVGGGIYTKAADVGADLVGKVVHGIP
EDDPRNPATIADNVGDNVGDVAGMGSDLFGSFAESTCAALVLGSSIGLSGGWDAMVFPVAVSAVGIFV
CLICSFIATDLRPVKEEKDVEMSLKVQLISTTVLMIPAVYLAAETFLPGEFMLKATVGNNVLTLHPWQ
AFLCVIMGAVGGLIIGLVTEYYTSHSYKPVRELADSCKTGAATNMIYGIALGYKSAIIPVLILAVVVY
GSFSMCDMYGVSLAAIGFLSNLATGLTIDVYGPVCDNAGGIAEMAELEPYVREKTDALDAAGNTTAAI
GKGFAIGSAALVSLALFGAFLFQDGVNMLSPLTFAFLIIGGMIPFAFAAMTMKSVGVAAMEMVLEVQR
QFDEKPHLLDANPTERPDYDACIGISTKASLKEMVAPGAMVILTPLLTGIFFGVYAVSGLLVGSLVAA
VQLAISMSNSGGAWDNAKKYIEKATPDSDLKGKGSDIHKAAVVGDTVGDPFKDTSGPALNIVMKLMAV
LSLVFADTFYATNGGAGVFQL

newV2.0.genewise.8.971.1 (TpVPP2)	Protein ID 158864

DNA sequence

ATGGCTGATATCATCCCCAATCTCATCTCCGAGGTCGCCTGCCTTGGCAAGGGCTTCCCTTGCGGAGA
AGATCCAATTACCCTCGTCTCTTTGAGTTTAGCATGCTCTGGTCTCGCTTTGGTTATGGTCTTCTATC
TCGTACACAAGGTGAGTTTTGTTGTGATGTGTATTGATGTTCATATGTTCATTGGCACCAATGCGAAC
TCATCAACTGCCACCTGACACCTCAACGCGTCCTTACATCTACAGCTCAAGCAAAACGAGCGTGGTAC
CGATGAAATGAACAAACTCTCCGACAAGATTCAATCTGGAGCCAAAGCCTTCCTCGTCACCGAGTACA
AATACCTCTCAATCTTTGTCCTACTCGTCTTTGCCGTCCTTGTCATCCTTTACTCTGTCGATCCACCC
ACCGGAGCTGGGAATGGAGATCGCGTCGACGGTATTCGTGCCGGTGCTTGCTTCTTGGCCGGTGCCAT
GCTCAGTGCGGGAGCTGGTTGGGCTGGTATGATCGTTGCTACAGATGCCAATGTTCGTACTACTCAAG
CTGCTGACAAGCAGGGTCTCGCCGTGGCTCTTCGTGTTGCCTTTACTGGAGGTGCTGTGATGGGATTC
ACCGTTGTTGGACTTGGATTGTTGGGAGTCTCCCTCTTCTTCTTCCTAATGACATTGAACCGATCTGG
TCAGTTCTTGGACAATGCTACCGATGCTGAGACCCTCATCGCTGCCATTGACTCTCTTGCAAGTTTCG
GATTTGGTGCCTCTTCCATTGCACTATTCGCACGTGTTGCCGGTGGTATCTACACAAAGGCTGCCGAT
GTTGGTGCTGATCTTGTTGGAAAGGTTGAGATGGACATTCCTGAAGATGATCCTCGTAATCCAGCTGT
GATTGCTGACAATGTGGGAGACAACGTTGGAGATGTTGCCGGAATGGGAGCTGATCTCTTCGAGTCCT
TCGTTGGATCTATCATTGCTGCGGCTACACTCGCGAACGGAGATCCTGCCAAGGTGGCCTTGCCATTC
TGGATCGCTGGAGCTGGAATTGTTGCCTCTGTATTTGGATTCTTCGCAGTAAGAACCAAGGATGGGGC
CAACCAACGTCAACTCATGATGGCTCTTCACAAGGGAGTTATCTTGTCGTCCTTCCTCGTTTTGGGTC
TCAGTGCCATCATCATTTGGAAGTTGTTTGAGGGAGATGCTTCAGATGAAGGATGGAAGCTGTGAGTA
ACATTGCGTGGCTGTTGGTTGTTTTCATTTGATCGCACAAAAGTCTAACATACTCGTCTTTGCAGCTA
CGCGTGCATTGCCATCGGCCTTATTGCTGGCATTCTTATTGGCCAGGCCACAGAATACTTTACTTCGT
ACTCCTTCTGGCCTACTCAGTCTATTACTCAAGCCGGAGTTACTGGTCCTGCCACCGTCATCATTCAA
GGAATCGGTGTAGGAATGATTTCGTGCGTTGCACCCGTTCTCGTCATTGTTGGAACCATCCTTGGGTG
TAATGCTTTGAGTGGACAGTACGGTATTGGTGAGTAGTTGTGTATCGTTGCTCATCTATTTGCTATCC
GTCGTTGCATAACTGTCTCTAACAAACTTTACTGGTGTTTATACTCAGCCATGTCTGCTGTTGGTATG
CTTAGTACCCTCGGCGTCACCCTTGCAACGGATGCCTACGGTCCTATCGCCGACAATGCTGGAGGAAT
CGCCGAGATGGCTGAACTTGAAGAACGTGTTCGTGAGACCACCGATGCACTTGACGCATTGGGCAACA
CCACTGCTGCAACAGGCAAGGGCTTCGCAATTGGATCAGCCGTGTTGACTGCCTTGTCATTGCTCAGT
GCTTTCAAGGATAAGGCTGCTGTGGAGACCGTTGACATCGGAGATCCCGTCGTGCTTAGCGGCGTCTT
GTTTGGAGCCATGCTTCCGTTCTTGTTCGCTGCTTTGACCATGCTTTCTGTGCAGAAGGCTGCTGGAG
CAATCATCATTGAAGTACGTCGTCAGTTCGCCCAGATTGAGGGATTGAGGGAGGGTACCGCCGAGGCC
GATTCCGACAAATGCGTTGCTATCTCCACTCAGTCATCTGTCGAGGAGATGATTCTTCCCGGTGTCTA
TGCCGTATTGAGTCCCCTCATGATTGGATTGTTGATTGGACCCAAGTGCCTCACTGGAATGTTGGGAG
GAGCAATTGCCTCTGGCATGATGCTTGCCCTCATGATGGCAAACGCAGGTGGTGCTTGGGATAACTCC
AAAAAGTACATTGAGATCGAAGGAGCTTGCGGAGGAAAGGGAACTACTACTCATAAAGCTTGTGTTGT
CGGAGATACTGTGGGAGATCCTTTCAAAGATACCAGCGGACCTTCTCTTAATATCCTCATCAAGCTCA
TGTCCATGATCTCTTTGACGATTGCGCCCATCATTCGCGGATGGGAAGACTGGCAATTCTGGTACTGG
GGATTCCTTCCTATCGGATTGGCCGCGATTGCAACTGTTCTCGTGTACCACTACTTCTGGAAGGATAC
ACAAGACATCTCTGCTCCTATCGAGGACGTCATCAAGACTGAGGCCGAGGACGTCATTAAGATTGAGG
CCGAGGACGTCAAGCAGGTTGACACTCCCGCTGCCGATGTGGAGGCACCTGCAGCTGAAGTGTCTGCC
GAGTAG

Protein (798 aa)

MADIIPNLISEVACLGKGFPCGEDPITLVSLSLACSGLALVMVFYLVHKLKQNERGTDEMNKLSDKIQ
SGAKAFLVTEYKYLSIFVLLVFAVLVILYSVDPPTGAGNGDRVDGIRAGACFLAGAMLSAGAGWAGMI
VATDANVRTTQAADKQGLAVALRVAFTGGAVMGFTVVGLGLLGVSLFFFLMTLNRSGQFLDNATDAET
LIAAIDSLASFGFGASSIALFARVAGGIYTKAADVGADLVGKVEMDIPEDDPRNPAVIADNVGDNVGD
VAGMGADLFESFVGSIIAAATLANGDPAKVALPFWIAGAGIVASVFGFFAVRTKDGANQRQLMMALHK
GVILSSFLVLGLSAIIIWKLFEGDASDEGWKLYACIAIGLIAGILIGQATEYFTSYSFWPTQSITQAG
VTGPATVIIQGIGVGMISCVAPVLVIVGTILGCNALSGQYGIAMSAVGMLSTLGVTLATDAYGPIADN
AGGIAEMAELEERVRETTDALDALGNTTAATGKGFAIGSAVLTALSLLSAFKDKAAVETVDIGDPVVL
SGVLFGAMLPFLFAALTMLSVQKAAGAIIIEVRRQFAQIEGLREGTAEADSDKCVAISTQSSVEEMIL
PGVYAVLSPLMIGLLIGPKCLTGMLGGAIASGMMLALMMANAGGAWDNSKKYIEIEGACGGKGTTTHK
ACVVGDTVGDPFKDTSGPSLNILIKLMSMISLTIAPIIRGWEDWQFWYWGFLPIGLAAIATVLVYHYF
WKDTQDISAPIEDVIKTEAEDVIKIEAEDVKQVDTPAADVEAPAAEVSAE

F.9 NCBI sequences

Acetabularia mediterranea	BAA83103
Arabidopsis thaliana AVP1	BAA32210
Arabidopsis thaliana AVP2	AAF31163
Arabidopsis thaliana AVP3	AAG09080
Beta vulgaris 1	AAA61609
Beta vulgaris 2	AAA61610
Carboxydothermus hydrogenoformans	ABB14908
Chara corallina	BAA36841
Chlamydomonas reinhardtii	CAC44451
Cucurbita moschata	BAA33149
Hordeum vulgare	BAA02717
Leishmania major	CAJ08309
Nicotiana tabacum TVP5	CAA54869
Nicotiana tabacum TVP9	CAA58701
Nicotiana tabacum TVP31	CAA58700
Oryza sativa 1	BAA08232
Oryza sativa 2	BAA31524
Plasmodium falciparum PfVP1	AAD17215
Plasmodium falciparum PfVP2	AAG21366
Pyrobaculum aerophilum	AAF01029
Rhodospirillum rubrum	AAC38615
Streptomyces coelicolor	BAD36743
Thermotoga maritima	AAD35267
Toxoplasma gondii	AAK38076
Triticum aestivum	AAP55210
Trypanosoma cruzi	AAF80381
Trypanosoma brucei	AAK95376
Vigna radiata	BAA23649.
Zea mays	CAG29369

Appendix G Aquaporin sequence details

Putative start sites (ATG) are shown in boxes. Red text indicates inferred amino acid sequence between putative translational start sites that differ from predicted gene models at JGI.

G.1 cDNA clone WS86

GTTTTTTTTTTTTTTTTTTTTTTAACTCGTTGAACCCCTATATTACACTTGCTTGGTGGCTACTACCGCT
CTATCTAATGAGGTAGAACCGACCGATGCTGTCAGCTCGCCTACACTAGCTCTGGCTGAGGCTGCGGGTG
GTGCATCTCCACTAAGACGCGGTACAGACCTGCGCCGCACACTCCGCCCAGAGAGTCGGCGACGAGCGGG
ATCCAGAAGTAGTAGTTTCGAATGGTGAAGACCTTGGAGCCCCAGCCGGCGATGGCCGTGAAGATGCGTG
GACCCAAGTCACGCGCAGGGTTCACAGCGTAGCCCGTGTTCATGCCGAAAGCCATGCCCAACGCCATGAT
CATGAGGCAGAAGGCAAAGGGAGAGCCCACTGGGCCCGCAGATCTATTGCGCTTGTCCGTGATGGCGTAA
ATACTGAGAACCAACATGGCAGTACCGATGAACTCGGTGTAGAAGGCTGTGTAGTTGGAGATGTTGTCGC
TCGGGTACGTGGAGAAGCTGCTCTGTGTGGTCTCGCGGTGCGGGTCGATGACGTTGAGGTTCTGGTACTG
CATGACGTAGATGATGAAGGCTCCGCAGAAGGCACCCAATAGTTGTGAGATCATGTAACCGGGAGCTTTC
CACCACGGCAGACGGCCGTACACGCAGTGCGCCAGTGTCACGGCCGTGTTGAGGTTGGCACCACCCTCGT
GCCGAATTCCTGCAGCCCGGGGGATCCACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCT

G.2 Genomic DNA - 9 kb contig

TGCCAGTTTTTCTTTTTATTAAAATTGCTGCATCCTCCCACCTACATCGTGTCCAGACGTGTGCATACCG
ATTAAAAGGCCATAACATCGCATCAACCTCACCAAGTTCGCGATGCCACAACTAGCTCGGTGAGAAAGTT
GTTCAGTGGGTCGCCTCTTGACGTCGACTTGCTTGCTTGGAGGCTCCACAGGATATCGCACCCGAAGCAG
AAAAACGCATTAGGTACACACCAAGGCACGCCACATTAGACCATACAACAACCAAGTCATCTCGATAGAA
ACGAACATAATCCGAACGCTTAGAAATTATCCGAAATTTAACAATCAAATAATCGCCGCTTCTTGTCGCT
GCAGCTTCCTCGTCACAGATCGGCGCTTGATACTGAAGAGAGGTGCCGTCAAACCACCGTGACAATGCTT
AAGTTACATCCGACCATTCTGCATTTTACCATCAACTGGTCAAGCAAAGCACAACAACATTCAACCAAAC
TCAAATTGACATGGCCCCACTGCCATCCGCAACCATGACCACACCCGACGTGGCTCCGGCCACAGCTACT
ACCGCCTCGCAGCCCTTTGACGACTTAGTCGACAAGGTCGAGAGCGGCTACGCCGCTATGCCTACCACCC
CAAAGACAGCCACGTCCTTCTCAACATCCAAACCGCAAGGTCTCCAGCGTTTCGCCGTGCAGAGCGTGCA
TCTACGCGAGTGCTTCGCCGAGTTCCTCGGCACCTTCGTCATGATCGTCTTCGGTATGGGCGTCAACAAC
CAAGTGACCAACTCGCAAGACGCCAACGGCACGTGGCTCAGCATCAACATGTGCTGGGGCATTGGTGTGC
TCATCGGCGTCTACTGCTCTGAGGGCATCAGTGGTGCCAACCTCAACACGGCCGTGACACTGGCGCACTG
CGTGTACGGCCGTCTGCCGTGGTGGAAAGCTCCCGGTTACATGATCTCACAACTATTGGGTGCCTTCTGC
GGAGCCTTCATCATCTACGTCATGCAGTACCAGAACCTCAACGTCATCGACCCGCACCGCGAGACCACAC
AGAGCAGCTTCTCCACGTACCCGAGCGACAACATCTCCAACTACACAGCCTTCTACACCGAGTTCATCGG
TACTGCCATGTTGGTTCTCAGTATCTACGCCATCACGGACAAGCGCAATAGATCTGCGGGCCCAGTGGGC
TCTCCCTTTGCCTTCTGCCTCATGATCATGGCGTTGGGCATGGCTTTCGGCATGAACACGGGCTACGCTG
TGAACCCTGCGCGTGACTTGGGTCCACGCATCTTCACGGCCATCGCCGGCTGGGGCTCCAAGGTCTTCAC
CATTCGAAACTACTACTTCTGGATCCCGCTCGTCGCCGACTCTCTGGGCGGAGTGTGCGGCGCAGGTCTG
TACCGCGTCTTAGTGGAGATGCACCACCCGCAGCCTCAGCCAGAGCTAGTGTAGGCGAGCTGACAGCATC
GGTCGGTTCTACCTCATTAGATAGAGCGGTAGTAGCCACCAAGCAAGTGTAATATAGGGGTTCAACGAGT
TAATTGTGTTTGTTCTAGATTTGTTCGTATGAATTCGAGTTGGTATTTTCTCAGGCTCGCTTGTGTCCAA
ACGAGGTACACACAGTATGTCACTGTATGTCTTACACTGTGTGGAACTTGAGCTGGCGTTTTCGCATTGT
ACTTCTGAAATGTGCTGTCAAAGCGTTGAAAACCAGTTTCGAGCATTCGTCTCTAACAGTAAGGTGCTAC
TACTATTAAAATCAACGTTTGTGTTCGCGATTGACAGTTGCAACTTTTACTAAAAAAAGCCTTTGACGTC
ATTTCCTTCCGTAGGCGTTTGAAGATCATCGGAATGTTGAAGATCTGCGCAGAACTTACCTGTCGTCGAC
TCAAACTGAGCGATACATGCTGTCGCTACGTAAATCGAAAGCGCTATGTTAAACCCGAAGTGTTTTGCGC
TGCGTGTTTCATTTTGCGGGTTCTTCCAAAACAAAAAAGTGAAGTCAACTGTCCTGCTTCCAATAGAGTT
GCTTTCGAAGAAAGGAGCATAACATCCTACGAAGGACCTCCATCGAAGCCATGACGGGAGCTCGAGGGCC

TTGTTCGAGCAATTGCTAATGTCTTGGATTGCAGTGTGCAACTGTCGAACTGGTTACAAGTACAGTTTTT
TTTAGTGATTATGATCGAGAAAAATCGTGTTTGCTCGCGGACGGATCAGGCACAAAAGAACCGCCTGAAT
GGCCGCAAAATTACGAGCCCATCGCAAACCGAATGGAAGTGGATCGTGTGCGATTCGCGTGGTCCGTGTC
AATGTGCTGCCAAAACCGCCGAATCGGCATGCCATCTGACACTGTATTCGTGTTAGACACCCTGAGCAAG
CAAAGCTATAGCGGACTCTGACTTGTGTTGACAGCACAATAAGATTCACTCACTGAGTGCGTGAGTCACT
CCGATTATTGAGAGACTCTCACCCACCAGTACTGTACGACCCTCCCCGGCGTGTCCTCTTGCTGCCGAAT
CACGACTCCCATATATCCATCTATGAAAGAGGTAACGCTTTGAGCACTGACTATTTGTTGGTATTAGCGC
TTAGCTCTTGGCGTATTGAACCACTCTTCGACAGTATTTGCACGACGACACAGTACTGTTGAAGTTACGG
AAAATCGACCGTTGTCCTCTTGTAACTTGCTACACGTAGGTGTACCGGTAGTTCGCTCGACAGTCTCAAA
CGAATCGATGGTAAATACTAGTCGAGCTAACAACCCTCTGTCATAGATGCGGCGCAGAGTCACCATGTTC
GCTTGCTCAGAAACTGAAGTGCTCATCCACATGCCACTCTCAACAATGCCTCGTCATTTCCAAGGGTTAA
TCTTAAGCTGCAGCGATACGGTCCCGTTGGCAAAACGTGATAGAATGCGAAGGCCGATAATAAACGCCCT
ACAGCACAACGCAGCTCCAACGTGTCGCGGGTGCCAGTTTCTAATCAAATATCTGTCTGCTTATTGGAGG
TAGGACTTTGAATCGGTCTATCGATCGGTTGGTGATGACTTGAAATTAGTCCAGCCGCGCTCCAATAAAC
GCTGTCGGAACTGTCACAATACGAGCCAGCTTCCAGCCGACAGCATGGGCTCGACGAGCATGTTTGCGCT
CGAGCTGGATTCGCCTATCTCTCGCCGCAAAAGAAACGCACCAAGCAAATCCAAATGCTCACCAAGCAGA
AAAAATCGCTATGAGTGATTAATCTACGATACGCCTTGTATTAAATATTAATACATAGTTAGTCGCAATT
GACCGCGACGGTCGCCTTCAGCCTTATTCTACATTTCACCACCAATTGATCAAGTCACCAACCACTCTCC
ACCTTCACCGTGACTAGTTAGCTGCATTTCTTGCATCCGACACTTCTACACAAGTCACCGGTTCTCCATA
CTGCACTCAACTACCATGACTTCCAAGCAAGACACCTCCATTCCCGTTTCGTACGACGAATTGGTCGACA
AGATCGAGACGGGTTACGTCGACGCACTGAACACACCGAAGCTCGCTTCTGCTTCGCTAGGCAAGCCTGT
CTCACCTTACGCCGTGCAGAGCGTGCACATGCGCGAGTGTTTGGCCGAGTTCCTCGGCACCTTTGTCATG
ATCATCTTCGGTATGGGCGTCAACAACCAGGTCGTGAACTCCGAGGAGAAGAACGGCACATGGCTCAGTA
TCAATATGTGTTGGGGCGTGGCCGTACTTATCGGTGTCTATTGCTCGGAAGGCATTAGTGGTGCCAACCT
GAACACCGCCGTGACACTGGCGCACTGCGTATATGGCCGTCTGCCGTGGTGGAAGGCACCGGGCTATATG
GCGTCCCAGTTGTTGGGAGCCTTCTGCGGAGCCTTCATCATCTACGTCATGCAGTACCAGAACCTCAACG
TCATCGACCCGTACTGCGAGACCACACAGAGCAGCTTCGCCACGTACCCGAGTGACAACATCTCCAACTA
CACAGCCTTCTACACCGAGTTCATCGGTACAGCTATGCTGGTGCTTGGCGTGTACGCTATCACGGATAGG
CGCAACAAGTCTGCAGGTCCCGTGGGCTCCCCCTTCGCCTTCTGTCTGCTCATCTGGGCTTTGGGTATGG
CGTTCGGCATGAACACGGGTTACGCCATCAACCCCGCTCGCGATTTCGGTCCTCGTCTCTTCACGTGCAT
GGCCGGGTGGGGCACCAAAGTGTTCGCGTTACGTAACTACTACTTCTGGATCCCGATCGTGGCTCCACTG
TGTGGCGGCGTGGCCGGCGCGGGTCTGTACCGCGTGATGGTGGAAATGCACCACCCGCAGCCACAACAAT
AAGTGTCAGCGTTGATTGAAAAGCACCCCGCCAGCAGGAGTTTAGACTGCAACTTCACGGAATAGAGTTC
ACATTTTGTTTTTAATGTTTTTTTTCAGGATCGCTCTTACGGCAAATACGTTTCAGTAGTTCCCTATTTT
ATGGTTGCATCCTCAGCCAACACATATCTAGGATTTACTTTTATTAATACCACTACAACTAGTATATTTG
ATACGTCATGCAGGTTTATCAATGGTATGGTGACGTGTGGAGTGAAAATGGTTGTCAATTCCTTGGTCGA
TGCAACCAGAGGATCTCCTACAGCTTTTAGTCAAGTTAACGTTAGGGTACAAAGAAATTAACAATGTAGG
AAGGAGAATCCTTAGGACTCCAATCCTATAATTGTAGTTTTTAGTTGCTGTATGCAAGTGTTCGAAGCAT
ACGAACAAGAACATCTATATATTAAACTCTGTTGACCCACTATCTCAGAGCTGGAGTGGCTGCGGAATGG
GCGAGTGGTGGATCTCCACAAGCAAGCGGTACAAGCCAGCACCAGCAACACCGCCCAAGAGGTCAGCCAC
GAGCGGGATCCAGAAGTAGTAGTTTCGAATGGTGAAGACCTTGGAGCCCCAGCCGGCGATGGCCGTGAAG
ATGCGTGGACCCAAGTCACGCGCAGGGTTCACAGCGTAGCCCGTGTTCATGCCGAAAGCCATGCCCAACG
CCATGATCATCAGACAGAAGGCAAAGGGAGAGCCCACTGGGCCCGCAGATCTATTGCGCTTGTCCGTGAT
GGCGTAGATACTGAGAACTAACATGGCAGTACCGATGAACTCGGTGTAGAAGGCTGTGTAGTTGGAGATG
TTGTCGCTCGGGTACGTGGCGAAGCTGCTCTGTGTGGTCTCACGGTTCGGGTCGATGACGTTGAGGTTCT
GGTACTGCATGACGTAGATGATGAAGGCTCCGCAGAAGGCACCCAATAGTTGTGAGATCATGTAACCGGG
AGCTTTCCACCACGGCAGACGGCCGTACACGCAGTGCGCCAGTGTCACGGCCGTGTTGAGGTTGGCACCA
CTGATGCCCTCAGAGCAGTAGACGCCGATGAGCACACCAATGCCCCAGCACATGTTGATGCTGAGCCACG
TGCCGTTGGCGTCTTGCGAGTTGGTCACTTGGTTGTTGACGCCCATACCGAAGACGATCATGACGAAGGT
GCCGAGGAACTCGGCGAAGCACTCGCGTAGATGCACGCTCTGCACGGCGAAGCGGGACATTCCCGTACGC
TCGGTGGGGACAAACGCGGTAGTCTTTGGACTTTCCGGTAGACCGCCGTAGCCGCTCTCGACCTTGTCGA
CCGCGTCGAAATCTTGGAGCGATTGGTTCATTAGCGCTCGAATGGTTTAATGCTTGCTATCCACAGAGGA
TATTGTAGAATGTGAGAATGGAAAACACTATGCCAGCAACAGCGCCAGACGAAAGAGAGAGCGCGAGTCG
TGGCACCCATGAGCCAACGTGCGCGTGAAGCGGTCAAAGATTAAAATTGTATGTTTGACTCCATAGCGCA
GGTGCTTGGACGCACGCCAAGCGCGTTTTTCTATAGCGCTTCAGAAAGTGCTCACCCTCCTTCAAATCGA
TCGCAGAGCCACGTCTTAACTTCGCATACACTACCATGAAGATGGTTTGACATGTGAGATTCACAGTGGT
CCTTCATGCTTTATAATTGACAAAACGAGCCCCTTTAGGCAGGAAACTGGAAACTCAGAAGTAGCTGCAT
GTATGCAAGAGCAAAAGTAAATCGCTCCATGATTAAAACCATTCCCAGTGACGTGGTGGTCTCGACTATG
ATGCATTTCAAGTTATCCCGTTGTGGCGAGAAATAGTTTTCCCAGGTAAAGAGTACTTACGACAAGTGGT
TTCGCTTAAATGCTGAGCCTGGAGCACAAGGATCATGGCGGGAGATGAGTCAGTGAATTGCAATAGAGAA
AACCGGGGACTCCGTGTCAGCTCAGTACAAATAAAGAGTTGACAAGGGACCCTGGTGTTTAACTTTAAAA

AAAAAAGAGTGTCTTTGGGCAAGCAACGACACAGCTGGTCTCTCCCATGTTTCATCCAGTCGAGCCGCGA
CTAGAAAAAAACAATCAGGTCAATGAGCGCTTGTCTCTTCGGCCCTGCCGGCTCAAACTCGGATGTATTC
ATTGGATTTTAGCTTTTGAACTTCAAATGCTTTGGGCCAGTAGTGTACGCTATCTAGTAGACGATACCCA
GCATTACTTTGGTAATGTTTGTCAATAAGTTGATGCTACATTTCCTAACGAACGTCGAGAACCGACTCCG
TTTGCTCAATCAGATGCTAAGTCTGCTTGGTGCCGCCGAAGCAGAAAAACGCAACAAAAGTTAACACATT
TAGAAATTCATTAGAAATGCAAATCAAAGGGTTAAATTCGGATACATCTTAAATTTGGAATCGTTAAATT
AGGTCCAAATTAAACTCGGTACCATATTCAAATCGACTCGACAAATAAACGAAACAAAAGTCTATGAGAT
GATAGTGGACGATTCGTCGATTGAGTTCAATCCCTCGTTAAGCCTAAAGACAAGGTCTATGATGCCAGAT
CAATTGCCATTCCTGTTTGCGAGCTGTTAGATGAGCAATCGTCTAAAATTGTTGCTCTGGATTGGTCGAG
GGCAGTACATGGTGTTGCTTGCTAAGGAAGTGCTCCATTGCAAAAATACGAGTCGTTTGTGATTCGATGA
TCAGCGCTGGCTTTCCATCATAGTTCAGTATCCAGAAAGCCGGCTTGTTGCGGCGTGGTTATCACTGAGA
CCGGGTCATACTTCTCTAGGCAAATATTCGAGGATGTAGGATCCGATGCGTACCAGGGTATTAAAAAACC
AATATCTGGTTATCACGGATTTCAATGTCTACACCAGCAAAAAGCTGCCAAGAAGATCAGTTATAGATTT
TCATTTGACGACCTCAAGCTAGAATTGGAGGTTATCCTCGTTTGACACATTACCGGCCTCAGCATTGCAA
AGATACGTGACACCAAACAGCACACATTGCCATGCAGGTCAGTCAACACCTCACTGTCAGTTGGCAACTT
ATTAACGCACCATATTAACGGTAAGTTATGCTGCCGAGTTGAAGCGACGACAAGTCTCCATGTGCGGGAT
CGGTCGAACGACTGGTGCTGCGGTTTCCGTGTCGCCGGCGCTTGATCGTTTTCTGAGTATTCGAGGTCTC
AAGCAAGACTTGGTCAACTTATATATCGAAATCAACGACGCAAACCAACTTCTCCTGTCTCATTTTGTAT
TTTACCACTTCCCCAGCCGATCGGTCCTCCTACTTTTACTATCCATGGCCCCGCTCCCATCTGCACAGGA
CCCAGTGGTCGAAGCTAACCCCGTCGTCACTGAGCCACTCGCCCTAAGCGACGACTTTGACAAAGCCGAG
AATGGCTACGCCGGTATGACTACAACCCCCAAGGAAGCAGTCGCGTACCCTGATCGAACAGGTATGGCCC
GCTTCGCCGTACAGAGCGTGCATCTACGCGAGTGCTTCGCCGAGTTCCTCGGCACCTTCGTCATGATCGT
CTTCGGTATGGGCGTCAACAACCAAGTGACCAACTCGCAAGACGCCAACGGCACGTGGCTCAGCATCAAC
ATGTGCTGGGGCATTGGTGTGCTCATCGGCGTCTACTGCTCTGAGGGCATCAGTGGTGCCAACCTCAACA
CGGCCGTGACACTGGCGCACTGCGTGTACGGCCGTCTGCCGTGGTGGAAAGCTCCCGGTTACATGATCTC
ACAACTATTGGGTGCCTTCTGCGGAGCCTTCATCATCTACGTCATGCAGTACCAGAACCTCAACGTCATC
GACCCGAACCGTGAGACCACACAGAGCAGCTTCTCCACGTACCCGAGCGATAACATCTCCAACTACACAG
CCTTCTACACCGAGTTCATCGGTACTGCCATGTTGGTTCTCAGTATCTACGCCATCACAGACCAGCGCAA
CCGTCCGGCTGGTCCCGTGGGCTCTGCATTCGCTTTCTGCCTCATGATCATGGCGTTGGGTATGGCGTTC
GGCATGAATACGGGTTACGCTGTGAACCCTGCTCGTGACTTGGGTCCACGTATCTTCACGGCCATCGCTG
GCTGGGGTTCGAAGGTGTTCACGACACGCAACTACTACTTCTGGATCCCGCTCGTGGCTGACCTCTTGGG
CGGTGTTGCTGGTGCTGGCTTGTACCGCTTGCTTGTGGAGATCCACCACCCGCCCCTCGCCCACCAGAAT
TAACTCTTGGAAACAAATACATTCGCATAGATAAATTCCTTCCTGTCAGTATACCATTAAAATGCTGCAC
ACAAGTTGACTTTCATGTGAACAAGGCGCTTGGTTGCAACATGTCCGCTCACGCTCGAGTTCTACTTTTG
CTTAGAAGATCGCAGCTTCAACCTGTCATGGTCATTTTACGAATAATCCTCATCTCTTGCAAATCTACAC
CTTTACTTTAAACCTATTTTTCCAAACTACATACACACTTCCCATTGAGCACTATCTAATAATTACAAAA
TTAGTCCAAATCCTTGTAAATTCCTGCCGTATCCGTTAGCCTGTCATTCCCAATTGAGACCCTCCCCATG
TCGGACCTGTCCAAGCTCGTTGTCACCTTCGCCGTGGGCTCAGCAGTGGGCGTCGCGGCCACGGCCCTGG
TCCTCCGTCAGCCCAAGAGCGTGAAGGCTCTGCCCAAGCTGTACATCTACGACCACTGCCCGTTCTGCGT
GCGTGCGCGTATGATCTTCGGCTTCAAGAAGGTCCCCCACGAGCTGGTATTCCTGGCCAACCACGATGAG
GCCACACCCATCGGCCTCGTGGGCTCTAAGCAGGCTCCGATCCTGCAACTAACCGATGGCCACGCTTTTC
CCGAGAGTATGGACATCGTCAAGTACGTGGACGAGCACTACGGCGGCCCGGTCGTGCTGGCCCCGTCCGC
GGACCGGCCCGAGATCAAGCAGTGGATCAAGGACACGGCCGACGTGTTCCGTCAGTTGTACCACCCACGA
TTCCACGCTGCTCCGTTCGCCGAGTTCGCCATGCTCGAGAGTCGCGAGTATTACCGCAAGAAGAAGGAGA
AAACCATTGGCCCGTTCCAGGACGCGCTGGACAAGACGCCGGAGCTCGTGGAGCAGGCCAATAAGTATCT
GGAACAGCTGGCCCCTATGTTTCACTCCAGCCACTCGGTCAATGAGAAGCTCAGCTACGACGACATCGAC
CTGTTCGGCCGTCTACGTGGTCTGACTCTCGTGCGCGATCTGGAGTGGCCCCCGAAGCTCCGCGAGTACG
TCGAC

G.3 Genomic DNA – Subclone pBS S1_S2

GTCGACATCTGAGGTTTCCTGCGTAGGTGTAGAAGTCGCATCTTCAGCCGCGTACTGGGCCATGGTGAAG
TGGTTGGCAGATTTAAGAGTCCGAAGTTGTTGGTTTGAGTACTTTAAAGCACAAATGAGAATGAGTCTTG
GAATGAACAGGCAAAAGTCGCTAATGCTCGAGCTTTCCATCTCTGCGTCTCTCGCGTGGGTTGCTGCATA
GGTTGGTAGCGACCAAGCAAGCAAGCGGATGCAGTGGGATGTCCAGCGCAGGCTGATCCAGGTGGAAATT
ATCGGCTACAAATTAACGTAAAAAAAAACTAGGAAAGCGTTATGCTTGACGGCACGTGCGTAGATTTTGA
GCGCGGTTCTCGGTGCTATCGTATTCTCGATTGCCTTGATTTCTCATGATGTGGATTTGGAGTGGAAGGT
AGTGGGGCCAAAGAAACGGATGCCAAATCTCTGTCTCATCTAGCATCTCACCCTACGCCACAGATCAATC
CCCGCTCCGATACTGGGGTAGCAATGGGATTCTCGATCTTTAGTATGATGCTGCCGTACTTGCAGTCAGT

GATCAAAATGGACAACAGAGCTAAGAGTGAGAAGTTAGACATGACTCTATGGCACCATTTTCCATTTGAA
CACATAGTATATAGGTAGTTATTTTATGAAGCTACGATCGACATTAAACTTAACAACCACCGACTACAAC
TTCGTAAGTACTTGGTCTGGGATCGGTCATACCGTTCTTAAGGTAGCGTCTGGCGCGGGTGGTGGATCTC
CACCAGCATGACGTACAGACCGGCACCAATCACGCCACCGATCAGTGGCCCCACGATCGGGATCCAGAAG
TAATGATCGCGCAGAGTGAAGACCTTGGAGCCCCAACCAGCGATACTCGTGAAGAGACGAGGGCCGAAGT
CACGCGCCGGGTTAATAGCGTAACCCGTGTTCATACCGATACACATACCGAGAGCCATGATCAAAAGGCA
GAAGGCGAAAGGAGCTCCCACTGGCCCAGCCGGACGGTTGCGCTGGTCCGTGATAGCGTAGATACCGAGT
AGCAACATGCCTGTACCGACTACTTCAGTGTAGAAGGCAGTATAGTTAGAAATATTATCGCTCGGGTATG
TAGCAAAGCTACTCTGCGTAGTCTCGCGGTCGGGATCAATGACGTTAAGGTTCTGCCACTGCATGACCCA
GATCACAAAGGCCCCCGCGAATGCACCCAAGACCTGTGAGATCATATATCCAGGTGCCTTCCACCACGGT
AGACGACCGTATACGCAGTGTGCTAGTGTCACTGCAGTGTTTAAGTGGGCTCCACTAATACCTTCCGTGC
AGTAGACGCCAATGAGCACGCCGATGCCCCAGCACATGTTGATACTGAGCCAAGTACCATTAGCGTCTTG
CGAGTTGGTCACCTGGTTGTTGACGCCCATACCGAACAAGATCATGACGAATGTGCCGAGAAACTCAGCG
AAGCATTCGCGCAGATGTACGCTCTGTACGGCGAAGCGAGATAGACTCGCGAGTCTGGCGTCAGCCGGGT
TGACGCTAGTTGTCTTAGGCGTCAACGGCATATCGATGTAGCTGACGTCGCGCTTGGCGTCGAGTTCTTC
GAACGGCTGCGGGGCTGGGTTGGCCGACACATCCGGGGTCTGGTTGGACGGGAACGGAGCCATTGTTGAG
GTTTGCGATGAAGGCTCGACCTTGGATGTGTTTGGAAGGCGTTAAAATGGAGAATGAGCGAGGCCTTCCG
AGCGAGGAAGTGTAATCGTCTCCACCAAAGGCAAGTTACCGTAGTGGAAGATGGCAGCGTCGGTGGGAGC
CAGAGCCCTGGCGTGAGAGCTGTTGAACAAGTTCAAGGCCGCTTGAGGCTACATTCACTTGACTTGACGG
TCAATTGGACACGAGTGAGCTCGGGTTACCTGTCTGACTCAGGTACTTCAGCTTGACCATTGCGTATCGC
GCGGTGACTCTTGCCAGGTGCGTAAACTGTGAAAGCAAGCAGTGGGACGTCTATCCGAGTGTATAAGTGT
TATTGCAATTTTTCGCCTCCGCTGCCGAGATGATAGCCCGGAGAAGGATAATGTAAGTGTTGAGTATGCA
TGTGTTGGTATCATGCCCGGATCGATGCTTTTGTCGCTCTGCTCCTTGAGCCAACACTAAGAGTATAAGC
ACAGACCTTACTCGTCTGCATCGATCGCGTTACCTTAACTGCTGTCGCCATTGTACAGTAGGGACTGCGA
GATCAGCGGAAGACGTGGCGTAGTTGCGACCAAGTTCAAGCGAGGCTGAATGTTGCGCACCAAGTTGTGC
GACTCACTAGTTGATCGAAACTTCAACGTTCAGGTCGTGAGCCAGTGCGTGCACACGAGGAATGCAATTG
CTCCCACGCCCGACGTATTGCAAGTATTAATCTGTACCAGCGTAGTTCCGGTTGCAGCACTGCAGGCGAT
GAACATTCTTGACTTCACATATCTTCAGCGTTTCTGCTTGGCGCTTTCTGAAGCGCCCCGCTCAGTGCCT
CAGTAATGAGTGCCTTTCTTAGTGGAATATGTGTCCCACAGTGCGTTGAAGATGTTTGACATTTAAATAG
CAGCATGCTCTTCAACTGTCAAACATGTTCTTCAACGGGCTGAAATCAGAAACGCAGCCTTCTGATTTGT
GTGCTTGACAAACAACCGGACCTTTTTTGGCGGATTAATCGTAAAAAAATGGTATCCTTTTGTTAATTTC
GGACCCCAAACGGAATTTCCCTATTTGAACAGTACAATATTAAAATTAGTCTCGTTTGCAGATTAAAAAA
TCAATTTTCAAAATTCAATAATTTCATTTACGAAAACTCACTGCAGCCATGTTTGCGTTTCTTGGCATGC
GCCAAAGTCGCGCGATGTTTTTCAGATGCCTTCCCTGACCATTCTCCTTTTTCAGCGCCCAGGCAAGGCT
TATCTGCCACGATTTTCACTCCAGCCAGATTGCTGTGGCTCACCGATCCTTCTGCTCTTGTAATGATGCC
CTATAGCTCCAAGTGTTTAAAGTTTTCCAGAAACCAACCAATGCTCAACATCCACCTAGGACTGCTGCAA
TACGGTATTGCCGACGTAATGGACGAGGACGGTGTGATCGTCGCACAACGAGTACAGGTGGTCCTTGTGA
TGCCTCCGCGACTTCGAAAGTGGCTGCACGCAGCCTCGCAGTATCTCCAGTCCAGATATCTCTACTCATA
CTGATAGGGACGATCTAGTCCCGCTAACGCCGTGGCATCATACCCTCCGCTAGCGCTGAGACTAACTCGC
AAGCTTATCGATACC

G.4 *P. nicotianae* aquaporins

PnAq1
DNA sequence
ATGGCTCCGTTCCCGTCCAACCAGACCCCGGATGTGTCGGCCAACCCAGCCCCGCAGCCGTTCGAAGA ACTCGACGCCAAGCGCGACGTCAGCTACATCGATATGCCGTTGACGCCTAAGACAACTAGCGTCAACC CGGCTGACGCCAGACTCGCGAGTCTATCTCGCTTCGCCGTACAGAGCGTACATCTGCGCGAATGCTTC GCTGAGTTTCTCGGCACATTCGTCATGATCTTGTTCGGTATGGGCGTCAACAACCAGGTGACCAACTC GCAAGACGCTAATGGTACTTGGCTCAGTATCAACATGTGCTGGGGCATCGGCGTGCTCATTGGCGTCT ACTGCACGGAAGGTATTAGTGGAGCCCACTTAAACACTGCAGTGACACTAGCACACTGCGTATACGGT CGTCTACCGTGGTGGAAGGCACCTGGATATATGATCTCACAGGTCTTGGGTGCATTCGCGGGGGCCTT TGTGATCTGGGTCATGCAGTGGCAGAACCTTAACGTCATTGATCCCGACCGCGAGACTACGCAGAGTA GCTTTGCTACATACCCGAGCGATAATATTTCTAACTATACTGCCTTCTACACTGAAGTAGTCGGTACA GGCATGTTGCTACTCGGTATCTACGCTATCACGGACCAGCGCAACCGTCCGGCTGGGCCAGTGGGAGC TCCTTTCGCCTTCTGCCTTTTGATCATGGCTCTCGGTATGTGTATCGGTATGAACACGGGTTACGCTA TTAACCCGGCGCGTGACTTCGGCCCTCGTCTCTTCACGAGTATCGCTGGTTGGGGCTCCAAGGTCTTC ACTCTGCGCGATCATTACTTCTGGATCCCGATCGTGGGGCCACTGATCGGTGGCGTGATTGGTGCCGG TCTGTACGTCATGCTGGTGGAGATCCACCACCCGCGCCAGACGCTACCT
Protein (311 aa)
MAPFPSNQTPDVSANPAPQPFEELDAKRDVSYIDMPLTPKTTSVNPADARLASLSRFAVQSVHLRECF AEFLGTFVMILFGMGVNNQVTNSQDANGTWLSINMCWGIGVLIGVYCTEGISGAHLNTAVTLAHCVYG RLPWWKAPGYMISQVLGAFAGAFVIWVMQWQNLNVIDPDRETTQSSFATYPSDNISNYTAFYTEVVGT GMLLLGIYAITDQRNRPAGPVGAPFAFCLLIMALGMCIGMNTGYAINPARDFGPRLFTSIAGWGSKVF TLRDHYFWIPIVGPLIGGVIGAGLYVMLVEIHHPRQTLP

PnAq2
DNA sequence
ATGGCCCCACTGCCATCCGCAACCATGACCACACCCGACGTGGCTCCGGCCACAGCTACTACCGCCTC GCAGCCCTTTGACGACTTAGTCGACAAGGTCGAGAGCGGCTACGCCGCTATGCCTACCACCCCAAAGA CAGCCACGTCCTTCTCAACATCCAAACCGCAAGGTCTCCAGCGTTTCGCCGTGCAGAGCGTGCATCTA CGCGAGTGCTTCGCCGAGTTCCTCGGCACCTTCGTCATGATCGTCTTCGGTATGGGCGTCAACAACCA AGTGACCAACTCGCAAGACGCCAACGGCACGTGGCTCAGCATCAACATGTGCTGGGGCATTGGTGTGC TCATCGGCGTCTACTGCTCTGAGGGCATCAGTGGTGCCAACCTCAACACGGCCGTGACACTGGCGCAC TGCGTGTACGGCCGTCTGCCGTGGTGGAAAGCTCCCGGTTACATGATCTCACAACTATTGGGTGCCTT CTGCGGAGCCTTCATCATCTACGTCATGCAGTACCAGAACCTCAACGTCATCGACCCGCACCGCGAGA CCACACAGAGCAGCTTCTCCACGTACCCGAGCGACAACATCTCCAACTACACAGCCTTCTACACCGAG TTCATCGGTACTGCCATGTTGGTTCTCAGTATCTACGCCATCACGGACAAGCGCAATAGATCTGCGGG CCCAGTGGGCTCTCCCTTTGCCTTCTGCCTCATGATCATGGCGTTGGGCATGGCTTTCGGCATGAACA CGGGCTACGCTGTGAACCCTGCGCGTGACTTGGGTCCACGCATCTTCACGGCCATCGCCGGCTGGGGC TCCAAGGTCTTCACCATTCGAAACTACTACTTCTGGATCCCGCTCGTCGCCGACTCTCTGGGCGGAGT GTGCGGCGCAGGTCTGTACCGCGTCTTAGTGGAGATGCACCACCCGCAGCCTCAGCCAGAGCTAGTGT AG
Protein (317 aa)
MAPLPSATMTTPDVAPATATTASQPFDDLVDKVESGYAAMPTTPKTATSFSTSKPQGLQRFAVQSVHL RECFAEFLGTFVMIVFGMGVNNQVTNSQDANGTWLSINMCWGIGVLIGVYCSEGISGANLNTAVTLAH CVYGRLPWWKAPGYMISQLLGAFCGAFIIYVMQYQNLNVIDPHRETTQSSFSTYPSDNISNYTAFYTE FIGTAMLVLSIYAITDKRNRSAGPVGSPFAFCLMIMALGMAFGMNTGYAVNPARDLGPRIFTAIAGWG SKVFTIRNYYFWIPLVADSLGGVCGAGLYRVLVEMHHPQPQPELV

PnAq3
DNA sequence
ATGACTTCCAAGCAAGACACCTCCATTCCCGTTTCGTACGACGAATTGGTCGACAAGATCGAGACGGG TTACGTCGACGCACTGAACACACCGAAGCTCGCTTCTGCTTCGCTAGGCAAGCCTGTCTCACCTTACG CCGTGCAGAGCGTGCACATGCGCGAGTGTTTGGCCGAGTTCCTCGGCACCTTTGTCATGATCATCTTC GGTATGGGCGTCAACAACCAGGTCGTGAACTCCGAGGAGAAGAACGGCACATGGCTCAGTATCAATAT GTGTTGGGGCGTGGCCGTACTTATCGGTGTCTATTGCTCGGAAGGCATTAGTGGTGCCAACCTGAACA CCGCCGTGACACTGGCGCACTGCGTATATGGCCGTCTGCCGTGGTGGAAGGCACCGGGCTATATGGCG TCCCAGTTGTTGGGAGCCTTCTGCGGAGCCTTCATCATCTACGTCATGCAGTACCAGAACCTCAACGT CATCGACCCGTACTGCGAGACCACACAGAGCAGCTTCGCCACGTACCCGAGTGACAACATCTCCAACT ACACAGCCTTCTACACCGAGTTCATCGGTACAGCTATGCTGGTGCTTGGCGTGTACGCTATCACGGAT AGGCGCAACAAGTCTGCAGGTCCCGTGGGCTCCCCCTTCGCCTTCTGTCTGCTCATCTGGGCTTTGGG TATGGCGTTCGGCATGAACACGGGTTACGCCATCAACCCCGCTCGCGATTTCGGTCCTCGTCTCTTCA CGTGCATGGCCGGGTGGGGCACCAAAGTGTTCGCGTTACGTAACTACTACTTCTGGATCCCGATCGTG GCTCCACTGTGTGGCGGCGTGGCCGGCGCGGGTCTGTACCGCGTGATGGTGGAAATGCACCACCCGCA GCCACAACAA
Protein (298 aa)
MTSKQDTSIPVSYDELVDKIETGYVDALNTPKLASASLGKPVSPYAVQSVHMRECLAEFLGTFVMIIF GMGVNNQVVNSEEKNGTWLSINMCWGVAVLIGVYCSEGISGANLNTAVTLAHCVYGRLPWWKAPGYMA SQLLGAFCGAFIIYVMQYQNLNVIDPYCETTQSSFATYPSDNISNYTAFYTEFIGTAMLVLGVYAITD RRNKSAGPVGSPFAFCLLIWALGMAFGMNTGYAINPARDFGPRLFTCMAGWGTKVFALRNYYFWIPIV APLCGGVAGAGLYRVMVEMHHPQPQQ

PnAq4
DNA sequence
ATGAACCAATCGCTCCAAGATTTCGACGCGGTCGACAAGGTCGAGAGCGGCTACGGCGGTCTACCGGA AAGTCCAAAGACTACCGCGTTTGTCCCCACCGAGCGTACGGGAATGTCCCGCTTCGCCGTGCAGAGCG TGCATCTACGCGAGTGCTTCGCCGAGTTCCTCGGCACCTTCGTCATGATCGTCTTCGGTATGGGCGTC AACAACCAAGTGACCAACTCGCAAGACGCCAACGGCACGTGGCTCAGCATCAACATGTGCTGGGGCAT TGGTGTGCTCATCGGCGTCTACTGCTCTGAGGGCATCAGTGGTGCCAACCTCAACACGGCCGTGACAC TGGCGCACTGCGTGTACGGCCGTCTGCCGTGGTGGAAAGCTCCCGGTTACATGATCTCACAACTATTG GGTGCCTTCTGCGGAGCCTTCATCATCTACGTCATGCAGTACCAGAACCTCAACGTCATCGACCCGAA CCGTGAGACCACACAGAGCAGCTTCGCCACGTACCCGAGCGACAACATCTCCAACTACACAGCCTTCT ACACCGAGTTCATCGGTACTGCCATGTTAGTTCTCAGTATCTACGCCATCACGGACAAGCGCAATAGA TCTGCGGGCCCAGTGGGCTCTCCCTTTGCCTTCTGTCTGATGATCATGGCGTTGGGCATGGCTTTCGG CATGAACACGGGCTACGCTGTGAACCCTGCGCGTGACTTGGGTCCACGCATCTTCACGGCCATCGCCG GCTGGGGCTCCAAGGTCTTCACCATTCGAAACTACTACTTCTGGATCCCGCTCGTGGCTGACCTCTTG GGCGGTGTTGCTGGTGCTGGCTTGTACCGCTTGCTTGTGGAGATCCACCACTCGCCCATTCCGCAGCC ACTCCAGCTC
Protein (298 aa)
MNQSLQDFDAVDKVESGYGGLPESPKTTAFVPTERTGMSRFAVQSVHLRECFAEFLGTFVMIVFGMGV NNQVTNSQDANGTWLSINMCWGIGVLIGVYCSEGISGANLNTAVTLAHCVYGRLPWWKAPGYMISQLL GAFCGAFIIYVMQYQNLNVIDPNRETTQSSFATYPSDNISNYTAFYTEFIGTAMLVLSIYAITDKRNR SAGPVGSPFAFCLMIMALGMAFGMNTGYAVNPARDLGPRIFTAIAGWGSKVFTIRNYYFWIPLVADLL GGVAGAGLYRLLVEIHHSPIPQPLQL

PnAq5

DNA sequence

ATGTGCGGGATCGGTCGAACGACTGGTGCTGCGGTTTCCGTGTCGCCGGCGCTTGATCGTTTTCTGAG
TATTCGAGGTCTCAAGCAAGACTTGGTCAACTTATATATCGAAATCAACGACGCAAACCAACTTCTCC
TGTCTCATTTTGTATTTTACCACTTCCCCAGCCGATCGGTCCTCCTACTTTTACTATCCATGGCCCCG
CTCCCATCTGCACAGGACCCAGTGGTCGAAGCTAACCCCGTCGTCACTGAGCCACTCGCCCTAAGCGA
CGACTTTGACAAAGCCGAGAATGGCTACGCCGGTATGACTACAACCCCCAAGGAAGCAGTCGCGTACC
CTGATCGAACAGGTATGGCCCGCTTCGCCGTACAGAGCGTGCATCTACGCGAGTGCTTCGCCGAGTTC
CTCGGCACCTTCGTCATGATCGTCTTCGGTATGGGCGTCAACAACCAAGTGACCAACTCGCAAGACGC
CAACGGCACGTGGCTCAGCATCAACATGTGCTGGGGCATTGGTGTGCTCATCGGCGTCTACTGCTCTG
AGGGCATCAGTGGTGCCAACCTCAACACGGCCGTGACACTGGCGCACTGCGTGTACGGCCGTCTGCCG
TGGTGGAAAGCTCCCGGTTACATGATCTCACAACTATTGGGTGCCTTCTGCGGAGCCTTCATCATCTA
CGTCATGCAGTACCAGAACCTCAACGTCATCGACCCGAACCGTGAGACCACACAGAGCAGCTTCTCCA
CGTACCCGAGCGATAACATCTCCAACTACACAGCCTTCTACACCGAGTTCATCGGTACTGCCATGTTG
GTTCTCAGTATCTACGCCATCACAGACCAGCGCAACCGTCCGGCTGGTCCCGTGGGCTCTGCATTCGC
TTTCTGCCTCATGATCATGGCGTTGGGTATGGCGTTCGGCATGAATACGGGTTACGCTGTGAACCCTG
CTCGTGACTTGGGTCCACGTATCTTCACGGCCATCGCTGGCTGGGGTTCGAAGGTGTTCACGACACGC
AACTACTACTTCTGGATCCCGCTCGTGGCTGACCTCTTGGGCGGTGTTGCTGGTGCTGGCTTGTACCG
CTTGCTTGTGGAGATCCACCACCCGCCCCTCGCCCACCAGAAT

Protein (377 aa)

MCGIGRTTGAAVSVSPALDRFLSIRGLKQDLVNLYIEINDANQLLLSHFVFYHFPSRSVLLLLLSMAP
LPSAQDPVVEANPVVTEPLALSDDFDKAENGYAGMTTTPKEAVAYPDRTGMARFAVQSVHLRECFAEF
LGTFVMIVFGMGVNNQVTNSQDANGTWLSINMCWGIGVLIGVYCSEGISGANLNTAVTLAHCVYGRLP
WWKAPGYMISQLLGAFCGAFIIYVMQYQNLNVIDPNRETTQSSFSTYPSDNISNYTAFYTEFIGTAML
VLSIYAITDQRNRPAGPVGSAFAFCLMIMALGMAFGMNTGYAVNPARDLGPRIFTAIAGWGSKVFTTR
NYYFWIPLVADLLGGVAGAGLYRLLVEIHHPPLAHQN

G.5 JGI DB analysis – *P. sojae*

estExt_Genewise1.C_950130 (PsAq1)	Protein ID 121749

DNA sequence

```
ATGTCGGGCAACGACGACAAGATCGAGAGCGGCTACGTCGGTGTGCAGGAACCAGACGTGTTCACGAA
CCTCGAGCCGGCCAAACCGTACCAAGTCAAGTCGCCGCTCTTCCGCGAGTGTCTCGCCGAGTTCCTGG
GCACCATGGTCATGATCATGTTCGGTGACGGCGTCGTGGCCCAGGTGGTGCTCAGCGAAGGCACCAAG
GGCGAGTACATCAACATCAACTTGTGCTGGGGCCTCGGTGTGCTCTTCGGTATCCACGTGTCGGGCGG
AGTTTCGGGCGCGCACCTGAACCCCGCGGTGACCACGACGCTCGCGTGGTTTGGCCGCATGGAGTGGA
AGAAGGTCCCGTACTACATCATCTCGCAGATCCTCGGCGCCTTCGCTGCCGCTTTCATCGTCTGGGTC
GTGTACTACCCCATGTTCAACGTGATCGACCCCCAGAAGACCTCGACGCAGGGCGTGTTTGCCACGTA
CCCGTACAGCAACGACGTGCCCGTTGGCACCTGCTTCCTCACGGAGGTGGTGGGCACGGCGATGCTCA
TGGGCTGCATCTTCGCTATCGGTGACGAGCTAAACAAGCCGGCCAACCCTTACACGCAGCCCGGCGCC
GTGGCCCTACTTGTGGTGGCCATCGGTATGGCTTTCGGCATGAACTCCGGCTACGCTATCAACCCCGC
GCGTGACTTCGGCCCTCGCCTCTTTTCGTGGATCGCTGGCTGGGGGTCCGAGGTGTTCACGCTCCGCG
ACAGCTACTTCTGGGTGCCGATTGTCGGCCCACTTCTCGGCGGCCCCATTGGCGCCGGTGTGTACGTT
GGCCTCATTGAGCACCACCACCCGCGCGAGTACAAGCACCCGCTCGAAGGTCAG
```

Protein (290 aa)

```
MSGNDDKIESGYVGVQEPDVFTNLEPAKPYQVKSPLFRECLAEFLGTMVMIMFGDGVVAQVVLSEGTK
GEYININLCWGLGVLFGIHVSGGVSGAHLNPAVTTTLAWFGRMEWKKVPYYIISQILGAFAAAFIVWV
VYYPMFNVIDPQKTSTQGVFATYPYSNDVPVGTCFLTEVVGTAMLMGCIFAIGDELNKPANPYTQPGA
VALLVVAIGMAFGMNSGYAINPARDFGPRLFSWIAGWGSEVFTLRDSYFWVPIVGPLLGGPIGAGVYV
GLIEHHHPREYKHPLEGQ
```

estExt_Genewise1.C_130061 (PsAq2)	Protein ID 113278

DNA sequence

```
ATGGTCGGAATGGCTCCTTCGATCCAAGACTACGACGCTGCCGACAAGGTCGAGAGCGGCTACGGCGG
TATGCCGGAATCCCCCAAGAACTCAACGTTCGTCTCAACTGAGCGCACCGGCCTGGCCCGCTTCGCCG
TCAAGAGCGTGCACATGCGCGAGTGCCTGGCCGAGTTCCTCGGCACCTTCGTCATGATCGTCTTCGGC
ATGGGCGTCAACAACCAAGTGACCAACTCGGAGGAGAAGAACGGCACGTGGCTCAGCATCAACATGTG
CTGGGGCATCGGCGTGCTCATCGGCGTCTACTGCTCCGAGGGCATCAGCGGCGCCAACCTTAACACGG
CTGTGACGTTGGCGCACTGCGTGTACGGCCGCCTGCCGTGGTGGAAGGCGCCGGGCTACATGATCTCG
CAGCTGCTGGGCGCCTTCATCGGCGCCTTCGTCATCTACGTCATGCAGTACCAGAACCTCAACGTCAT
CGACCCGCAGCGCGAGACCACGCAGAGCAGCTTCTCCACGTACCCTAGCGACAACATCTCCAACTACA
CGGCCTTCTACACGGAGTTCGTGGGCACGGCCATGCTCGTGCTCAGCATCTACGCCATCACGGACAAG
CGCAACCGCGCGGCCGGCCCCGTGGGCGCCGCGTTCGCCTTCTGCCTCATGATCATGGCGCTCGGCAT
GGCCTTCGGCATGAACACGGGCTACGCCGTGAACCCTGCGCGCGACTTCGGTCCTCGCCTGTTCACGT
TCTGCGCCGGCTGGGGCTCCAAGGTCTTCACGACGCGCAACTACTACTTCTGGATCCCGATCGTGGCG
GACCTGATGGGCGGCGTTGCTGGCGCTGGTTTGTACCGCCTGCTCGTGGAGATCCACCACCCGCCGCT
CCTGCACCAGAAC
```

Protein (299 aa)

```
MVGMAPSIQDYDAADKVESGYGGMPESPKNSTFVSTERTGLARFAVKSVHMRECLAEFLGTFVMIVFG
MGVNNQVTNSEEKNGTWLSINMCWGIGVLIGVYCSEGISGANLNTAVTLAHCVYGRLPWWKAPGYMIS
QLLGAFIGAFVIYVMQYQNLNVIDPQRETTQSSFSTYPSDNISNYTAFYTEFVGTAMLVLSIYAITDK
RNRAAGPVGAAFAFCLMIMALGMAFGMNTGYAVNPARDFGPRLFTFCAGWGSKVFTTRNYYFWIPIVA
DLMGGVAGAGLYRLLVEIHHPPLLHQN
```

estExt_fgenesh1_pm.C_860008 (PsAq4)	Protein ID 109301

DNA sequence

ATGCGTCCAAGTCCCCAAGCCGACTGCAACACCGTCCAAGAGCCTTTGGAGAAGGGCAGCAAGTCTTT
CTCCGCCTTCGCCAGTACGCCGTACACGCCGCACGAAGCTGTTGACCCATGCAGACGCTCCCTGCACC
CTACGAACCGCAGTATGAACGACCGCGAGGAAGATCGCCACAAGCCGTGGTTCGTGACCAAGAACCCG
CACCTCCGTGAATGCTACTCGGAATTCCTCGGCACCTTCGTCATGATCGCCTTCGGCATGGGCGTCAA
CAACCAAGTCGTGCTCTCCGAAGAGAAAGAGGGGACCTGGCTCAGTATCAACATGGCGTGGGGGATCG
CCGTCCTCATGGGCGTCTACTGCTCCGAAGGAGTGAGCGGCGCGCACATGAACCCGTCCGTGACGCTG
GCCCACGCCGTCTACGGCCGTCTCGCGTGGCGCAAGGTGCCGGGCTACGTGCTGTCGCAGTTCCTCGG
AGCGTTCGTGGGCGCCGTCGCCATCTACCTGCTCGACTACCAGCGTCTGTCGAAGGCCGACCCGGACA
AGGAGACCATGTACCACAACTTCGCGACGCACCCGAACCCGGAGATCAGCAACCTCACGGCGTTCTAC
ACGGAGGCGCTGGCCACGGGTATGCTGCTGCTGTGCGTGTACGCCATCACGGACCAACGCAACCGCTC
GCCGGGCACGGTCGGCACGCCGTTCGCGTTCGCTCTGATGATCATGGCGCTGGGCATGAGCTTCGGCA
TGAACACGGGGTACGCGATGAATCCGGCGCGCGACTTCGCTCCTCGTTTGTTCACGTTCTTCGCGGGG
TACGGGTCGAAGGTGTTCACGGAGAACGGTTGCTACTTCTTGATCCCGATGTTCGCTCCGCTGATCGG
CGGCGTGCTGGGCGCTGGTGCGTACGAGTTGCTTGTGCAGGTGCAGCACCCGCACGACCCAGCGGAGT
AC

Protein (318 aa)

MRPSPQADCNTVQEPLEKGSKSFSAFASTPYTPHEAVDPCRRSLHPTNRSMNDREEDRHKPWFVTKNP
HLRECYSEFLGTFVMIAFGMGVNNQVVLSEEKEGTWLSINMAWGIAVLMGVYCSEGVSGAHMNPSVTL
AHAVYGRLAWRKVPGYVLSQFLGAFVGAVAIYLLDYQRLSKADPDKETMYHNFATHPNPEISNLTAFY
TEALATGMLLLCVYAITDQRNRSPGTVGTPFAFALMIMALGMSFGMNTGYAMNPARDFAPRLFTFFAG
YGSKVFTENGCYFLIpmFAPLIGGVLGAGAYELLVQVQHPHDPAEY

estExt_fgenesh1_pm.C_4760001 (PsAq5)	Protein ID 109702

DNA sequence

ATGGCTTCCAAGCAAGACACGAACATCGCCGCCGTCTCCTACGACGAGCTGGTCGACAAGGTCGAGAC
CGGATACGTCGACGCGCCGAGCACGCCCAAGGCCGGCTCCGCCTCGCACGGCAAGGCCCCGTCGCGCT
ACGCCGTGCAGAGCGTGCACATGCGCGAGTGCCTGGCCGAGTTCCTCGGCACCTTCGTCATGATCGTC
TTCGGCATGGGCGTGAACAACCAGGTCGTCAACTCCGAGGAGAAGAACGGCACGTGGCTCAGCATCAA
CATGTGCTGGGGCGTGGCCGTGCTCATCGGCGTCTACTGCTCCGAGGGCATCAGCGGCGCCAACCTCA
ACACGGCTGTGACGTTGGCGCACTGCGTCTACGGCCGCCTGCCGTGGTGGAAGGCGCCGGGCTACATG
GTCTCGCAGGTGCTGGGCGCCTTCTGCGGCGCCTTCGTCATCTACGTCATGCAGTACCAGAACTTGAA
CGTCGTCGACCCGAACCGCGAGACGACGCAGGGAAGCTGGTCCACGTACCCTAGCGACAACATCTCCA
ACTACACGGCCTTCTACACGGAGTTCGTGGGCACGGCCATGCTGGTGCTGGGCGTGTACGCCATCACG
GACAAGCGCAACAAGTCGGCGGGCCCCGTCGGCTCTCCGTTCGCCTTCTGCCTGCTCATCTGGGCGCT
CGGCATGGCCTTCGGCATGAACACGGGCTACGCCATCAACCCTGCGCGCGACTTCGGCCCGCGTCTGT
TCACTTGTTTGGCGGGCTGGGGCACCAAGGTCTTCACGTTGCGCAACCACTACTTCTGGATCCCCATC
GTGGCTCCGCTGTGCGGCGGCGTGGCGGGCGCGGGGCTGTACCGCGTCATGGTGGAGATGCACCACCC
GCAGCCGCAGCAG

Protein (299 aa)

MASKQDTNIAAVSYDELVDKVETGYVDAPSTPKAGSASHGKAPSRYAVQSVHMRECLAEFLGTFVMIV
FGMGVNNQVVNSEEKNGTWLSINMCWGVAVLIGVYCSEGISGANLNTAVTLAHCVYGRLPWWKAPGYM
VSQVLGAFCGAFVIYVMQYQNLNVVDPNRETTQGSWSTYPSDNISNYTAFYTEFVGTAMLVLGVYAIT
DKRNKSAGPVGSPFAFCLLIWALGMAFGMNTGYAINPARDFGPRLFTCLAGWGTKVFTLRNHYFWIPI
VAPLCGGVAGAGLYRVMVEMHHPQPQQ

estExt_fgenesh1_pg.C_4760002 (PsAq6)	Protein ID 145096

DNA sequence

```
ATGGCTCCGACCCCGTCCAACCAGACCGCCGGCGTGTCGCCCGCCTGCCAAGTCCAGGGCTTCGACGA
CCTGGCCGACAAGGTCGAGAGCGGCTACGCCGGCATGCCCACCACGCCCAAGGCCACCACTCCGCTCT
ACCCCAACGAGGACACGCCGCGGGGCCTCGCCCGCTTCGCCGTCGAGAGCGTGCACCTGCGCGAGTGC
TTCGCCGAGTTCCTCGGCACCTTCGTCATGATCCTCTTCGGCATGGGCGTTAACAACCAAGTGACCAA
CTCGGAGCTCGAGAAGGGCACGTGGCTCAGCATCAACATTTGCTGGGGCATCGGCGTCATGATCGGCG
TCTACTGCTCCGAAGGCATTAGCGGCGCCAACCTCAACACCGCCGTCACGCTCGCGCACTGCGTCTAC
GGCCGCCTGCCGTGGTGGAAGGCTCCGGGTTACATGATCTCGCAGCTCCTGGGGGCCTTCTGCGGCGC
CTTTTGCATTTGGGTCATGCAGTGGCAGAACCTGAACGAGATCGACCCCAACCGCGAGACCACGCAGA
GCAGCTTCGCCACGTACCCGCGGGACAACATCTCCAACTACACGGCCTTCTACACCGAGTTCGTCGGC
ACGGCCATGCTCGTCATGAGCATCTACGCCATCACGGACACCAAGAACCGCCCCGCCGGCAAGTACGG
CAACGCCTTCGCGTTCGCCCTCATGATCATGGCGCTCGGCATGGCCTTCGGCATGAACACGGGCTACG
CCGTCAACCCGGCGCGTGACCTGGGCCCGCGCATCTTCACGGCCATCGCCGGCTGGGGATCGAAGGTC
TTCACGCTGCGTCACTACTACTTCTGGATCCCGATCGTGGCCGATTCGCTGGGCGGCGTGTGCGGCGC
GGGTCTGTACCGCCTGTTAGTTGAGATCCACCACCCGCGGGGCCCGCTGCTT
```

Protein (312 aa)

```
MAPTPSNQTAGVSPACQVQGFDDLADKVESGYAGMPTTPKATTPLYPNEDTPRGLARFAVESVHLREC
FAEFLGTFVMILFGMGVNNQVTNSELEKGTWLSINICWGIGVMIGVYCSEGISGANLNTAVTLAHCVY
GRLPWWKAPGYMISQLLGAFCGAFCIWVMQWQNLNEIDPNRETTQSSFATYPRDNISNYTAFYTEFVG
TAMLVMSIYAITDTKNRPAGKYGNAFAFALMIMALGMAFGMNTGYAVNPARDLGPRIFTAIAGWGSKV
FTLRHYYFWIPIVADSLGGVCGAGLYRLLVEIHHPRGPLL
```

estExt_fgenesh1_pm.C_180005 (PsAq7)	Protein ID 108598

DNA sequence

```
ATGGCAAGAAACTCCAGAAGCTCGAGGAACTCGGATCTGGAGAAAGCTGAGCACGTCTACTTCGACGT
CGGCACCCCTGACCCGATCAACCACCTGCAGAACCTGACCCCGGTCGAGAACGTGCCGTTCATAGTCA
AGAACCAGTTGGTCAAGGCCATGATGGCTGAGTTCCTCGCCATGTTCGTCACCATGCTCTTCGGGCTG
TGCTGCATGCTGCAGACCGTTCTAAGCAGCGGAACCGACGGCAGCTTCGTGACCATCGCGCTCTGTTG
GGGCCTCGCCTTCTTCTTCGGCATCTCCATCGCCGGCGGTGTCTCGGGCTCCCACCTGAATCCGGCCA
TCACGACGACGCTGGCGCTGGCGAAGATGCTGCCGTGGAAGAAGGTACCGTTCTACATCTTGAGCCAG
ATCCTCGGGTCATACGCCGCTGCGTTCTTCGCTTACGTCCTGTATCGCCCGATGCTCAACGAGTTGGA
CCCGGACCGGCTGACGACTCACACCATCTTCGCCACATACCCGCACGAAAACGTGGGCAACTTCACGT
GTTTCGCGACGGAATTCGTGGCGACGGCACTGCTGGTGATTGGGATCCTGGCATTGCTGGACCAGCAC
AACCGACCGATTGGAAGGAAGGCCGCTCCGGCTGCAATTGGGGCCCTGGTGAGCACGCTCGCCATGGG
GTTCGGTATGAACTCGGGGCTGGCCATGAACCCTGCGAAGGACTTGGGCCCTCGGTTGTTCATTTGGA
GTGCTGGCTGGGGCTCCCGCGTGTTCTCACGCAGTAATTACTTCTTCTGGATTCCCATTGTTGCGCCA
ACCTTGGGCGCTGCTGTCGGCGGATTCATTTACGAGGGTATGGTCGGCTACCACCACGTCGACAAGAA
CCGTGGTCCGCCTCCGGAGTACCGATTC
```

Protein (304 aa)

```
MARNSRSSRNSDLEKAEHVYFDVGTPDPINHLQNLTPVENVPFIVKNQLVKAMMAEFLAMFVTMLFGL
CCMLQTVLSSGTDGSFVTIALCWGLAFFFGISIAGGVSGSHLNPAITTTLALAKMLPWKKVPFYILSQ
ILGSYAAAFFAYVLYRPMLNELDPDRLTTHTIFATYPHENVGNFTCFATEFVATALLVIGILALLDQH
NRPIGRKAAPAAIGALVSTLAMGFGMNSGLAMNPAKDLGPRLFIWSAGWGSRVFSRSNYFFWIPIVAP
TLGAAVGGFIYEGMVGYHHVDKNRGPPPEYRF
```

estExt_fgenesh1_pg.C_860081 (PsAq9)	Protein ID 140543
DNA sequence	
ATGGCTCCGTTCCCGTCCAACCAGACCCAGGACGTCCAAGTGCCGCAGCCGTTCGAGGACGTCGACGG CAAGCACGAGGCCAGCTACAACGACATGCCGTTGACGCCCAAGGCCGCCGCCCCCGCTAAGCTGCAGC GGCTGTCTCGCTTCGCCGTACAGAGCGTGCATCTGCGCGAGTGCTTCGCCGAGTTCTTGGGCACCTTC GTCATGATCCTCTTCGGCATGGGCGTCAACAACCAAGTGACCAACTCGGAGGAGAAGAACGGCACGTG GCTCAGCATCAACATGTGCTGGGGCATCGGCGTGCTCATCGGTGTGTACTGCACGGAAGGCATCAGTG GTGCACACCTGAACACGGCCGTGACACTCGCGCACTGCGTCTATGGTCGTCTGCCGTGGTGGAAGGCT CCGGGCTACATGATCTCACAGGTTCTTGGGGCCTTCGTCGGTGCCTTTGTCATCTACGTCATGCAGTA CCAGAACCTGAACGTCATCGACCCGAACCGCGAGACCACGCAGAGCAGCTTCTCCACGTACCCGAGCG ACAACATCTCCAACTACACGGCCTTCTACACTGAAGTGGTCGGTACGGCCATGCTGCTGCTCGGCATC TACGCCATTACGGACCAGAAGAACCGCCCGGCCGGACCTGTGGGCGCTCCGTTCGCCTTCTGCCTGTT GATCATGGCTCTGGGCATGTGCATTGGCATGAACACGGGCTACGCCATCAACCCGGCGCGCGACTTCG GCCCTCGCCTGTTCACGAGCATCGCTGGCTGGGGCTCCAAGGTCTTCACCCTGCGCGACCACTACTTC TGGATCCCCATCGTGGGCCCGCTGATCGGCGGTGTTATCGGCGCCGGACTGTACATTCTGCTCGTGGA GATCCACCACCCGCCGCAGACGCTCCCT	
Protein (304 aa)	
MAPFPSNQTQDVQVPQPFEDVDGKHEASYNDMPLTPKAAAPAKLQRLSRFAVQSVHLRECFAEFLGTF VMILFGMGVNNQVTNSEEKNGTWLSINMCWGIGVLIGVYCTEGISGAHLNTAVTLAHCVYGRLPWWKA PGYMISQVLGAFVGAFVIYVMQYQNLNVIDPNRETTQSSFSTYPSDNISNYTAFYTEVVGTAMLLLGI YAITDQKNRPAGPVGAPFAFCLLIMALGMCIGMNTGYAINPARDFGPRLFTSIAGWGSKVFTLRDHYF WIPIVGPLIGGVIGAGLYILLVEIHHPPQTLP	

estExt_fgenesh1_pg.C_950050 (PsAq10)	Protein ID 141202
DNA sequence	
ATGGATCCGCTTTCATACAGCTACAAGGTCGACGAGGACTACACCCCCGTTCAGTCTCCTCGACCGCT CGAGCTCACCGTCCCTGCCACTCGCCCTTACCAGCTCAAGTCTCCACTGCTACGAGAGTGTCTCGCAG AATTCCTCGGGACATTCGTCCTCGTGCTCTTCGGAGACGGAGCCGTCGCTCAAGTCACGCTCAGCGAC AACGCCAAAGGGAATTACACCACGCTGTGCCTCGGGTGGGGAGTGGGCATCCTCTTCGGCATCCACGT CTCGGGAGGCGTCTCCGGCGCGCACATCAACCCCGCCATCACCACCACGCTCGCGCTTTTCCGACGTT TTGAATGGCGCAAGGTCGTCCCCTACGTCGTGGCGCAGACGCTCGGCGCCTTCGTGGCCGCCTTCCTC ATATGGGCCGTCTACAGGCCGCTCTTCGACGTTATCGACCCGCACAAGACGACCACGCAGGGCGTCTT CGCCACGTATCCCTACAGCAGCGACGTGTCCGTGGGCACTTGCTTCTTGACGGAGATAGTGGGCACCG CTCTGCTGCTGTGCGGACTCTTTGCCATCGGCGACGAGCTCAATAAACCCGCCAACCCATACTCGCAG CCCGGCGCAGTGGCACTGCTGGTCGTGGGTATTGGCATGGCGTTCGGCATGAACTCGGGTTTCGCGCT GAACCCCGCGCGCGACTTGGGGCCTCGGTTGTTCGCCTGGGTGGCCGGGTGGGGCTCTCGGGTGTTCA CGCTGCGTGACGCGTACGGTTGGGTCCCGCTGGTAGCTCCATTGCTGGGCGGCGCGATTGGAGCGGGC GTGTACGTTGGCCTCATCGAGCACCACCACCCCCGCCAGTACAGCCAAGAGCACAACGCCCAGTTCCC TGACGTGATGGAGCGAATCGATCTCTTTAGTACTTCATCGTACAAGCCG	
Protein (311 aa)	
MDPLSYSYKVDEDYTPVQSPRPLELTVPATRPYQLKSPLLRECLAEFLGTFVLVLFGDGAVAQVTLSD NAKGNYTTLCLGWGVGILFGIHVSGGVSGAHINPAITTTLALFRRFEWRKVVPYVVAQTLGAFVAAFL IWAVYRPLFDVIDPHKTTTQGVFATYPYSSDVSVGTCFLTEIVGTALLLCGLFAIGDELNKPANPYSQ PGAVALLVVGIGMAFGMNSGFALNPARDLGPRLFAWVAGWGSRVFTLRDAYGWVPLVAPLLGGAIGAG VYVGLIEHHHPRQYSQEHNAQFPDVMERIDLFSTSSYKP	

estExt_fgenesh1_pm.C_10018 (PsAq11)	Protein ID 108153

DNA sequence

ATGCCTTCGGACGCCTTCAACGAGAAGCACGAGCTCGGCTACGTCGACCTGGAAGCTCACGATGGCCA
CTCCACCACTACGGGCACGCACGTCCTCCTCGAGGAAGTGGCCGTTGTTCCGTACCAAGTCAAGAGCC
AATTCGCCAAGGAGATGATGGCCGAGTTCCTCGCCACCTTCATCTGCATGCTCTTCGGCCTGTCGTGC
ATGGCCCAAGTCACTCTGAGCGGAGGAAGCGCCGGCAGCTTCGTGACCATCGCATTGGCCTGGGGGTT
CGCCTACTTCATCGGCATCTCGATCGGTGGCGGCGTATCTGGCGCCCACCTGAACCCAACGGTGACCG
TGACGCTCGCTCTGCTCAAGATGCTGCCGTGGAAGAAGGTGCCGTTCTACATCCTCATTCAGACTGTC
GCCGCCTACGTCTCGGCCCTGGTGGTCTACATCGTGTATCGTCCGCTGTTTAATGAGGTGGACCCGGA
CCGCACCACGACCCACACAGTCTTCGCGACGTTCCCGCACGAAAACGTTGGCAACTTCACGTGCTTCC
TGACCGAGTTCCTCGCCGCAGCTCTGCTGATCCTTGGGATTCTGGCGCTGCTGGACCAGCACAACCGC
CCCATTGGCCGTCACGCGGTCCCGCCGGCCATCGGCGTGCTCGTGAGCGCTATCGCTATGGGCTTTGC
CACGAACACAGGCCTGGCCGCCAACGCTGCTCGCGACCTGGGTCCTCGTCTCTTCATGCTTACTGCAG
GCTGGGGATCGCGCGTGTTCACACTCGGTAATTATTACTTCTGGGTCCCGCTCGTGGCTCCGATCCTG
GGCGGCGCTGCCGGAGCATTCGTGTACGAGGCCATGGTGGGCTACCACCACCCTGGACGCTCGTTCGG
ACCGTACCCGGAGTACCGATTC

Protein (302 aa)

MPSDAFNEKHELGYVDLEAHDGHSTTTGTHVLLEEVAVVPYQVKSQFAKEMMAEFLATFICMLFGLSC
MAQVTLSGGSAGSFVTIALAWGFAYFIGISIGGGVSGAHLNPTVTVTLALLKMLPWKKVPFYILIQTV
AAYVSALVVYIVYRPLFNEVDPDRTTTHTVFATFPHENVGNFTCFLTEFLAAALLILGILALLDQHNR
PIGRHAVPPAIGVLVSAIAMGFATNTGLAANAARDLGPRLFMLTAGWGSRVFTLGNYYFWVPLVAPIL
GGAAGAFVYEAMVGYHHPGRSFGPYPEYRF

estExt_fgenesh1_pm.C_4180001 (PsAq12)	Protein ID 109697

DNA sequence

ATGGCCACGCTCCCATCCACCCAGAACCCCGACGTTGAGGCCGTCCCCGTCGTCGCAGAGCCCCTGGC
CTTCAACGACGACTTCGACAAGACTGAGAATGGTTATGCTGGCATGACGACCACACCCAAGGGCGCTG
CCGCCTACACCGACCGCCAGGGCTTGGCCCACCTCGCCGTCAAGAGTGTGCACATGCGCGAGTGCTTC
GCCGAGTTCCTCGGCACCTTCGTCATGATCGTCTTCGGCATGGGCGTCAACAACCAAGTGACCAACTC
GGAGGAGAAGAACGGCACGTGGCTCAGCATCAACATGTGCTGGGGCATCGGCGTGCTCATCGGCGTCT
ACTGCTCCGAGGGCATCAGCGGCGCCAACCTCAACACGGCTGTGACGTTGGCGCACTGCGTGTACGGC
CGCCTGCCGTGGTGGAAGGCGCCGGGCTACATGATCTCGCAGCTGCTGGGCGCCTTCATCGGCGCCTT
CGTCATCTACGTCATGCAGTACCAGAACCTCAACGTCATCGACCCGCAGCGCGAGACCACGCAGAGCA
GCTTCTCCACGTACCCTAGCGACAACATCTCCAACTACACGGCCTTCTACACGGAGTTCGTGGGCACG
GCCATGCTCGTGCTCAGCATCTACGCCATCACGGACAAGCGCAACCGCGCGGCCGGCCCCGTGGGCGC
CGCGTTCGCCTTCTGCCTCATGATCATGGCGCTCGGCATGGCCTTCGGCATGAACACGGGCTACGCCG
TGAACCCTGCGCGCGACTTCGGTCCTCGCCTGTTCACGTTCTGCGCCGGCTGGGGCTCCAAGGTCTTC
ACGACGCGCAACTACTACTTCTGGATCCCGATCGTGGCGGACCTGATGGGCGGCGTTGCTGGCGCTGG
TTTGTACCGCCTGCTCGTGGAGATCCACCACCCGCCGCTCCTGCACCAGAAC

Protein (312 aa)

MATLPSTQNPDVEAVPVVAEPLAFNDDFDKTENGYAGMTTTPKGAAAYTDRQGLAHLAVKSVHMRECF
AEFLGTFVMIVFGMGVNNQVTNSEEKNGTWLSINMCWGIGVLIGVYCSEGISGANLNTAVTLAHCVYG
RLPWWKAPGYMISQLLGAFIGAFVIYVMQYQNLNVIDPQRETTQSSFSTYPSDNISNYTAFYTEFVGT
AMLVLSIYAITDKRNRAAGPVGAAFAFCLMIMALGMAFGMNTGYAVNPARDFGPRLFTFCAGWGSKVF
TTRNYYFWIPIVADLMGGVAGAGLYRLLVEIHHPPLLHQN

estExt_fgenesh1_pm.C_950004 (PsAq13)	Protein ID 109359

DNA sequence

ATGCATCGCGAAAGCGCGGACAAGGTTGACGCGATGATGCGAGTCAGCAGCATGCCCAAGATGGACGT
GAACTATATGGGCGAACCGGACGTGGCCATGATGTCGACGGGGCCGGCCCGCAAAGGAATGTTCCAAA
CCAAGTCTCCTCTCTTCCGGGAATGCATGGCCGAGTTCCTCGGCATGTTCGTCTTCATGCTGTTCGGC
ACCGGCGTCGTCGCCCAAGTCGTGCTGAGCGAAGGCACGAAAGGCGAGTTCATCAGCATCAACTTCTG
CTGGGGTCTCGGCATCCTGTTCGGTATCCACGTCTGCGGTGGTGTATCGGGCGCGCACTTGAACCCGG
CGGTGACTATGACGCTGGCCCTGTTCGGCCGCTTCGAGTGGAAGAAGGTCCCGTGCTACGTGATAGCT
CAAATGCTGGCGGCGTTCCTGGCCGCGGCCATTGTCGGCATGGTGTACGACCCGCTGATCAAAGTCTC
CGACCCGCACATGAAGACGACCCAGGGAATCTTCGCCACGTATCCGTACAGCAATGACGTCCCCATTG
GCACCTGCTTCATGACGGAGGTGGTGGGCACGGCGCTGCTGGTCGGCTGCCTCTTCGCCATCGGCGAC
GAAATGAACAAGCCGGCCAGCCCGTATAGCCAACCCGGAGCTGTCGCGCTACTGGTGGTGGCCATTGG
CATGGCCTTTGGAATGAACACCGGCTACGCCATCAACCCGGCGCGCGACCTCGGTCCTCGCTTCTTTA
CTCTGTGTGCTGGTTGGGGGCCCAAGGTGTTCACGCTAAGCGAGTCGTACTGGTGGGTGCCGATCGTC
GGACCGATGCTTGGCGGCCCGATCGGCGCCGGCCTGTACGTGGCCCTGGTGGAGCAGCACCACCCCGT
CGAGTTCGATCACACGCACTCCTAAGCGCATGCAGCTGCGTCCTTGTCGTTGATAAAGTGCAGTAAAA
GTTGCGTTCGATGAGTCTCAAATCGTCTTGTGTATTTCGTTGGTCTCTCAAAACTCCGCTCTGGAGGT
GGTAACCGAGGCCGTCGGCAAGATGTTGCAGGCAAGAAGACTGGATTTACACGCTTCCTAACGCCGTG
ACACCACTTCGCGTTGAAGTT

Protein (302 aa)

MHRESADKVDAMMRVSSMPKMDVNYMGEPDVAMMSTGPARKGMFQTKSPLFRECMAEFLGMFVFMLFG
TGVVAQVVLSEGTKGEFISINFCWGLGILFGIHVCGGVSGAHLNPAVTMTLALFGRFEWKKVPCYVIA
QMLAAFLAAAIVGMVYDPLIKVSDPHMKTTQGIFATYPYSNDVPIGTCFMTEVVGTALLVGCLFAIGD
EMNKPASPYSQPGAVALLVVAIGMAFGMNTGYAINPARDLGPRFFTLCAGWGPKVFTLSESYWWVPIV
GPMLGGPIGAGLYVALVEQHHPVEFDHTHS

estExt_fgenesh1_pm.C_100023 (PsAq14)	Protein ID 108464

DNA sequence

ATGACGCGAACCTCGGACTTGGAGAAGAACAGCCTCACCTACGCCGACCTAGAGTCCCCGGACCCGAC
CCACCACCTGCAGCACCTGACGCCGGTCGAAAACGTGCCGTACATGGTCAAGAGTCAATTCGCCAAGG
AGATGATGGCCGAGTTCCTCGCCACCTTCGTCACCATGCTCTTCGGGCTGTCCTGCATGACGCAAGTG
GTGCTGAGCAGTGAGGCCAGCGGCAACTTCGTGACCATTGCGCTGTGCTGGGGTCTCGCCTTCTTCTT
CGGCATCACCATCGGCGGCGGCGTCTCCGGAGCTCACTTGAACCCGGCGGTGACCACGACGCTGGCGC
TGCTCAAGCTGCTGCCGTGGAAGAAGGTGCCGTTCTACATCCTCAACCAAGTCGTCGCCGCCTACGTG
GCCGCGCTCTTCGTGTACATTTTGTACCGGCCGATGTTCAACGAGGTGGATCCGGACCGAACCACGAC
ACACACCATCTTCGCGACGTACCCGCACGAGAACGTGGGCAACTTCACGTGCTTCTTGACGGAGTTCG
TGGCGACCGCTCTGCTGATCCTGGGCATTTTGGCGCTGCTGGACCAGCACAACCGTCCCATTGGAAAG
AAGGCCGTGCCGCCTGCCGTCGGAGCATTGGTGAGCACAATCGCCATGGGGTTCGCCATGAACACGGG
GCTGGCCATCAACCCTGCTCGGGACCTGGGTCCTCGTCTCTTCATGCTGTGCGCTGGCTGGGGTTCGC
GCGTTTTCTCTCTAAGTCAGTACTACTTCTGGATCCCCATTGTGGCTCCGATCACGGGCGGCGCTGCC
GGTGCGTTCGTGTACGAGGCCCTCATCGGATACCACCACGTGGAGAAGAACCGTGGCCCGCACCCGGA
GTACCGATTC

Protein (298 aa)

MTRTSDLEKNSLTYADLESPDPTHHLQHLTPVENVPYMVKSQFAKEMMAEFLATFVTMLFGLSCMTQV
VLSSEASGNFVTIALCWGLAFFFGITIGGGVSGAHLNPAVTTTLALLKLLPWKKVPFYILNQVVAAYV
AALFVYILYRPMFNEVDPDRTTTHTIFATYPHENVGNFTCFLTEFVATALLILGILALLDQHNRPIGK
KAVPPAVGALVSTIAMGFAMNTGLAINPARDLGPRLFMLCAGWGSRVFSLSQYYFWIPIVAPITGGAA
GAFVYEALIGYHHVEKNRGPHPEYRF

estExt_fgenesh1_pm.C_130007 (PsAq15)	Protein ID 108516

DNA sequence

ATGGCACACTTCAACCTGGAGCCAAGCAATGGCCCCGCGCCGCCGGAGACCGGCATGACGACGGCGCC
GTACTCGGCCTTCCACCTGCCCAACTTCGCGCACCGCATCATCGACCGCTGCGAGCGCTCGCTGCACC
ACACGAACCGCAGCCAGTACGAGCGCGACGAGGAGCACCACAAGAAGTGGTTCATCATGAACCGCGCG
CACATCCGCGAGTGCCTGGCCGAGTTCCTCGGCACCTTCGTCATGATCTGCTTCGGCATGGGCGTCAA
CAACCAGGTGTCGCTCTCGGCCGAGGCCAACGGCACGTGGCTCAGCATCAACATGGCCTGGGGCATCG
GCGTGCTCATGGGCGTCTACTGCTCTGAGGGCGTCAGCGGCGCGCACTTGAACTGCGCCGTCACCTTT
GCCCACGCCGTGTACGGCCGCCTGCCATGGTGGAAGGTGCCGGGCTACGTGGTCTCGCAAGTTGTTGG
CGCCTTCCTGGGCGCGCTCGCCATCTACCTGCTGGACTACCAGAAGCTCAACAAGGTGGACCCGGACC
AGGAGACGACGCAGAGCAACTTCGCCACGTACCCGAGCGCCGACATCAGCAACATTACGGCGTTCTAC
ACGGAGGCGCTGGCTACGGCTATGCTGCTGCTGTGCATCTACGCCATCACGGACCAGAACAACCGCTC
TCCCGGGACTGTGGGTACGCCGTTTGCCTTTGCTCTGATGATCATGGCGCTCGGTATGAGCTTCGGCA
TGAACACGGGCTACGCGATGAACCCTGCGCGAGACTTTGGGCCTCGCCTGCTGACGTACGTGGTGGGC
TACGGCTCGAAGGTCTGGACTGCGGATAGTTACTACTTCTGGATCCCGATCTGCGGGCCGTTGCTTGG
GGGTGTTATTGGCGCGGGACTGTACACTTTCCTGGTGCAGATGCAACACCCGCATGAACACGCGGCT

Protein (317 aa)

MAHFNLEPSNGPAPPETGMTTAPYSAFHLPNFAHRIIDRCERSLHHTNRSQYERDEEHHKKWFIMNRA
HIRECLAEFLGTFVMICFGMGVNNQVSLSAEANGTWLSINMAWGIGVLMGVYCSEGVSGAHLNCAVTF
AHAVYGRLPWWKVPGYVVSQVVGAFLGALAIYLLDYQKLNKVDPDQETTQSNFATYPSADISNITAFY
TEALATAMLLLCIYAITDQNNRSPGTVGTPFAFALMIMALGMSFGMNTGYAMNPARDFGPRLLTYVVG
YGSKVWTADSYYFWIPICGPLLGGVIGAGLYTFLVQMQHPHEHAA

estExt_fgenesh1_pm.C_860010 (PsAq16)	Protein ID 109303

DNA sequence

ATGGCCCCACTCCCCTCCACGCAAAGGCCCGACGTGGCGTCTCCGAGTCGAGCTCCGCAGCCGTTCGA
CGACCTCGCGGACAAGGTCGAGAACGGGTACGCGAGCATGCCGCTGACCCCCAAGGCCACGACCAACT
TCGCGCCTGAGGCCAAGCGCACCGGCCTGGCCCGCTTCGCCGTCAAGAGCGTGCACATGCGCGAGTGC
CTGGCCGAGTTCCTCGGCACCTTCGTCATGATCGTCTTCGGCATGGGCGTCAACAACCAAGTGACCAA
CTCGGAGGAGAAGAACGGCACGTGGCTCAGCATCAACATGTGCTGGGGCATCGGCGTGCTCATCGGCG
TCTACTGCTCCGAGGGCATCAGCGGCGCCAACCTTAACACGGCTGTGACGTTGGCGCACTGCGTGTAC
GGCCGCCTGCCGTGGTGGAAGGCGCCGGGCTACATGATCTCGCAGCTGCTGGGCGCCTTCATCGGCGC
CTTCGTCATCTACGTCATGCAGTACCAGAACCTCAACGTCATCGACCCGCAGCGCGAGACCACGCAAA
GCAGCTTCTCCACGTACCCTAGCGACAACATCTCCAACTACACGGCCTTCTACACGGAGTTCGTGGGC
ACGGCCATGCTCGTGCTCAGCATCTACGCCATCACGGACAAGCGCAACCGCGCGGCCGGCCCCGTGGG
CGCCGCGTTCGCCTTCTGCCTCATGATCATGGCGCTCGGCATGGCCTTCGGCATGAACACGGGCTACG
CCGTGAACCCTGCGCGCGACTTCGGTCCTCGCCTGTTCACGTTCTGCGCCGGCTGGGGCTCCAAGGTC
TTCACGACGCGCAACTACTACTTCTGGATCCCGATCGTGGCGGACCTGATGGGCGGCGTTGCTGGCGC
GGGGCTGTACCGCCTGCTCGTGGAGATCCACCACCCGCCGCTCCCGCACCAGAAC

Protein (313 aa)

MAPLPSTQRPDVASPSRAPQPFDDLADKVENGYASMPLTPKATTNFAPEAKRTGLARFAVKSVHMREC
LAEFLGTFVMIVFGMGVNNQVTNSEEKNGTWLSINMCWGIGVLIGVYCSEGISGANLNTAVTLAHCVY
GRLPWWKAPGYMISQLLGAFIGAFVIYVMQYQNLNVIDPQRETTQSSFSTYPSDNISNYTAFYTEFVG
TAMLVLSIYAITDKRNRAAGPVGAAFAFCLMIMALGMAFGMNTGYAVNPARDFGPRLFTFCAGWGSKV
FTTRNYYFWIPIVADLMGGVAGAGLYRLLVEIHHPPLPHQN

estExt_fgenesh1_pg.C_950046 (PsAq17)	Protein ID 141198

DNA sequence

```
ATGGCGACCCCTTCGAACGCAGCGGACAAGATCGAGCACGGCTACATGGACCTGGAAGCCCACGAGAC
GGCTGTGAACGGCGGGGCCCCCGTGAACACGAGCCGCTTCGCCATCAAGAGCCCGGCCCTTCGAGAGT
TCTGCGCCGAGTTCGTGGCCGTCTTCGTCATGATGAGCTTCGGCCTGGCCGCGACGTGCCAGTCCATC
CTGAGCAGCAGCAAGTACGGCGACTTCACGCAGATCGTCTTCGGCTGGGGTATCGGAGTGCTCTTCGG
CATCCACATCGCCGGCGGCGTCTCCGGCTCGCACCTGAACCCGGCCATCACGACCACGGCGGCGCTCT
TCGGCATGTTCCCGTGGAAGAAGGTGCCGACCTACATTGCCGCGCAGACGCTGGCCTCGTACCTGGCC
GCTCTGTTCGTGTACGTGCTCTACCGGCCGCTGCTCAACGTGGCCGACCCGGAGAGGCAGACTACGCA
CGCCATCTTCGCGACCTACCCGAACGAGTACGTGCCCACCTCCGTGGCCTTCCTGACCGAGATGTTCG
CGACGGCGCTGCTGGTGGGCGGTATCTTCGCCGTGCTGGACCAGCACAACCGCCCGGCCAGCCCGTTC
AGCGCCCCGAGCGCCGTGGCGCTGCTGGTGGTCGGTATCGGCATGGCCTTCAGCGTCAACACCGGCTG
CGCCATCAACCCCGCTCGCGACTTCGGCCCTCGCCTCATGATGCTGACGGCCGGCTGGGGAACACGAG
TTTTCTCGCTCGATAGTTACTACTTTTGGGTCCCCATTGTTGCCCCCACCGTCGGCGGCGCGCTCGGC
GGGCTTGTGTACGTGGTGCTCGTGGGCCACCACCACCCGCAAAAGGTCACCGTCGTCCGCAAGCACGG
CGAGCAGTTCCACTTC
```

Protein (300 aa)

```
MATPSNAADKIEHGYMDLEAHETAVNGGAPVNTSRFAIKSPALREFCAEFVAVFVMMSFGLAATCQSI
LSSSKYGDFTQIVFGWGIGVLFGIHIAGGVSGSHLNPAITTTAALFGMFPWKKVPTYIAAQTLASYLA
ALFVYVLYRPLLNVADPERQTTHAIFATYPNEYVPTSVAFLTEMFATALLVGGIFAVLDQHNRPASPF
SAPSAVALLVVGIGMAFSVNTGCAINPARDFGPRLMMLTAGWGTRVFSLDSYYFWVPIVAPTVGGALG
GLVYVVLVGHHHPQKVTVVRKHGEQFHF
```

estExt_fgenesh1_pg.C_950049 (PsAq18)	Protein ID 141201

DNA sequence

```
ATGGCCCGCTTGTCCGACAGCGGCAACAAGGACAAAGACGGCTACGTGGGCCTGCAGGAGGCCGACGT
GGCCCTGACGCCGGCATCGGCTGTGTCCAAGCCGTACCAGGTCAAGTCGCCGCTGCTGCGCGAGTGCA
TGGCCGAGTTCATCGGCACCATGGTGCTCATCATGTTCGGCGACGGCGTCGTGGCCCAGGTGGTGCTC
AGCGAGAGCACCAAGGGCGAGTACATCAACATCAACTTGTGCTGGGGCCTCGGAGTTCTCTTCGGCAT
CCACGCATCGGGCGGTGTTTCGGGTGCACACCTGAACCCGGCCGTCACTACGACGCTCGCTCTCTTCG
GCCGCTTCGAGTGGCGTAAAGTGATCCCGTACATCATCGCGCAGGTGTTGGGTGCGTTCGTGGCGGCC
TTCATCGTCTGGGCCGTCTACTACCCCATGTTCAACACCATCGACCCGAACAAGGAGACCACGCAGGG
TGTCTTCTCCACGTACCCGTACAGTGATGACGTCCCGGTCGGCACCTGCTTCCTCACTGAGGTGGTGG
GCACTGCATTGCTGCTGGGTGGTATCTTCGCTCTGGGTGACGAGCTCAACAAGCCGGCCAGCCCGTAC
ACGCAGCCGAGCGCTGTGGCCCTGCTGGTTGTAGCGATTGGCATGGCGTTCGGCATGAACTCGGGCTA
CGCCATCAACCCGGCCCGTGACTTCGGCCCGCGGTTGTTCTCGCTCTGTGCTGGCTGGGGCTCGCGCG
TGTTCACCCTGCGGGACCACTACTTCTGGGTCCCGATCGTCGGCCCGCTGCTCGGAGGTGCCATCGGT
GGCGGCGTGTACGTTGGCCTGGTCGAGCACCACCACCCGCGCGACTACAAGCACCCGCTCGACAAC
```

Protein (294 aa)

```
MARLSDSGNKDKDGYVGLQEADVALTPASAVSKPYQVKSPLLRECMAEFIGTMVLIMFGDGVVAQVVL
SESTKGEYININLCWGLGVLFGIHASGGVSGAHLNPAVTTTLALFGRFEWRKVIPYIIAQVLGAFVAA
FIVWAVYYPMFNTIDPNKETTQGVFSTYPYSDDVPVGTCFLTEVVGTALLLGGIFALGDELNKPASPY
TQPSAVALLVVAIGMAFGMNSGYAINPARDFGPRLFSLCAGWGSRVFTLRDHYFWVPIVGPLLGGAIG
GGVYVGLVEHHHPRDYKHPLDN
```

estExt_fgenesh1_pm.C_950010 (PsAq21)	Protein ID 109365

DNA sequence

ATGGCCGAGTACAAACCTGGATACGGAGGCTACGCCGACCTCGAAGGCAACGGCCGCTTGGCGCCCTA
CGCCATTCGCAGCCAGGATATGCGCGCCTACATGGCCGAGTTCGTCGGCACCTTCATCCTCGTGCTCA
TCGGCGATGGCTCCGTGGCCCAGTTCGTGCTCAGCAAGAGAGCCGCCGGTGACTACCTCTCCGTGAAC
CTGTGCTGGGGCATCGCGCTGCTGTTCGGCGTCCACTTCTCCGGTGGCGTGAGCGGTGGCCACCTCAA
TCCTGCGGTCAGTGTGACGATGGCCCTCTTCAAGCGCTTCGAGTGGCGCAAGGTCCCGGGCTACATCA
TCGCGCAGACGTTGGGCGCCTTCGTTGCTGCACTGGTGCTCTACATCGTCTACTACCCGTGGCTCGAC
ATCGTGGATCCGGAGCGCGAGTTCACGCAGGGCATCTTCGCCACGTACCCGAACCCACAGATCCCCAA
CTGGACCGCGTTCGCCAACGAGGTCATCGGCACGGCCCTGCTCGTTGGTGGCATCTTCGCTCTGTGTG
ACCAGATCAACAAGCCCGCGAGTCCCTACAGCTTCCCCGGCGCAGTGGGCCTCCTACTCACAGGTATC
GGCATGTCGTTCGGCCTCAACACAGGCTACGCTCTCAACCCCGCCCGTGACTTTGGCCCTCGCCTCCT
TACGCTCTTCGGCGGCTGGGGCTGGAAGGTGTTCTCGTCGCACGGCGGATACTTCTGGATCCCGATCC
TGGGCCCGTTCATGGGCGCTGTCCTGGGTGCGGCCATGTACGTGGGCCTCGTGGAGCTCCACCACCCG
CCTCAG

Protein (274 aa)

MAEYKPGYGGYADLEGNGRLAPYAIRSQDMRAYMAEFVGTFILVLIGDGSVAQFVLSKRAAGDYLSVN
LCWGIALLFGVHFSGGVSGGHLNPAVSVTMALFKRFEWRKVPGYIIAQTLGAFVAALVLYIVYYPWLD
IVDPEREFTQGIFATYPNPQIPNWTAFANEVIGTALLVGGIFALCDQINKPASPYSFPGAVGLLLTGI
GMSFGLNTGYALNPARDFGPRLLTLFGGWGWKVFSSHGGYFWIPILGPFMGAVLGAAMYVGLVELHHP
PQ

estExt_fgenesh1_pg.C_860085 (PsAq22)	Protein ID 140547

DNA sequence

ATGGCTCCACTGCCGTCCGCAGCAATGGCCACGCCCAACGTGGCATCTATCGCTGCCACCACCGCTGC
CGCTCAGCCGTCCGACGACGCGGCCGACAAGGTCGAGTGCGGCTACGCCGTGATGCCCACGACCCCCA
AGGCCATGCCGTTCGCCGCCCTCGCCTCCAGCGGCGCAAAGCCGCAGGGCCTGGCCCGCTTCGCCGTG
CAGAGCGTGCACGTGCGCGAGTGCTTCGCCGAGTTCCTCGGCACCTTCGTCATGATCGTCTTCGGCAT
GGGCGTCAACAACCAAGTGACCAACTCGCAGGACGCCAACGGCACGTGGCTCAGCATCAACATGTGCT
GGGGCATCGGCGTGCTCATCGGCGTGTACTGCTCCGAGGGCATCAGCGGCGCCAACCTCAACACGGCT
GTGACGTTGGCGCACTGCGTCTACGGCCGCCTGCCGTGGTGGAAGGCGCCGGGCTACATGCTGTCGCA
GCTGCTGGGCGCCTTCTGCGGCGCCTTCGTCATCTACGTCATGCAGTACCAGAACCTCAACGTCATCG
ACCCGCAGCGCGAGACTACGCAGAGCAGCTTCGCCACGTACCCGCGGGACAACATCTCCAACTACACG
GCCTTCTACACGGAGTTCGTGGGCACGGCCATGCTGGTGCTCAGCATCTACGCCATCACGGACAAGCG
CAACCGCGCGGCCGGCCCCGTGGGCGCACCCTTCGCCTTCGCCCTCATGATCATGGCGCTCGGCATGG
CCTTCGGCATGAACACGGGCTACGCCGTCAACCCGGCGCGTGACCTGGGCCCGCGCCTCTTCACGGCC
ATCGCCGGCTGGGGATCGAAGGTCTTCACGCTGCGTCACTACTACTTCTGGATCCCGATCGTGGCCGA
TTCGCTGGGCGGCGTGTGCGGCGCGGGTCTGTACCGCCTGTTAGTTGAGATCCACCACCCGCGGGGCC
CGCTGCTT

Protein (320 aa)

MAPLPSAAMATPNVASIAATTAAAQPSDDAADKVECGYAVMPTTPKAMPFAALASSGAKPQGLARFAV
QSVHVRECFAEFLGTFVMIVFGMGVNNQVTNSQDANGTWLSINMCWGIGVLIGVYCSEGISGANLNTA
VTLAHCVYGRLPWWKAPGYMLSQLLGAFCGAFVIYVMQYQNLNVIDPQRETTQSSFATYPRDNISNYT
AFYTEFVGTAMLVLSIYAITDKRNRAAGPVGAPFAFALMIMALGMAFGMNTGYAVNPARDLGPRLFTA
IAGWGSKVFTLRHYYFWIPIVADSLGGVCGAGLYRLLVEIHHPRGPLL

estExt_fgenesh1_pm.C_10016 (PsAq23)	Protein ID 108151
DNA sequence	
ATGTCAGCGGACGCCTACAACGACAAACACGGCGACGAGTACATCAACCTGGAGGCTCACGACGGACT CTCCACCGACGTGGGCTCGCTCAACGGCGTGACCCACGTTCCGTACCAAGTCAAGAGCGCCTTCGCCA GGCAGATGATGGCCGAGTTCTTCGCCACCTTCATCTGCGTCGTCGTCGGGCTGGCCTGCACAGCGCAG GTCACGCTGAGCTCGGGCACTGCAGGCAGCTTCGTGACCGTCGCGCTCGCGTGGGGCTTCGCCTACTT CCTCGGCATCACAGTCGGCGGAGGCGTGTCCGGCGCCAACCTGAACCCGACCGTGACCGTCGCGCTCG CGCTGCTCGGCATGCTGCCGTGGAAGAAGGTGCCCTTCTACATCCTCGCGCAGACTGTCGCGTCCTAC GTGGCCGCGATGGTCGTCTACATCCTGTACCGCCCCATGCTCAACATCGTGGACCCCGAGCGCACGTC GACCAACGCCATCTTTGCCACGTACCCACACGAGAACGTGGGCAACTTCACGTGCTTCCTGACCGAGT TCTTCGCCGCGGCGCTGCTGGTCGTGGGCTTCCTGGCGCTGCTCGACGAGCACAACCGCCCCATCGGC CGCGCCGCCGTTCCCGCCGCCATCGGCGTGCTCGTCAGCGCCATCGCCATGGGCTTCGGCATGAACAC GGGCCTGGCCATCAACGCCGCCCGCGACCTCGGCCCTCGCCTCTTGATGCTCACGGTCGGCTTCGGTT CACGCGTGTTCTCGCTCAGCCACTACTACTTCTGGGTCCCGCTCGTGGCCCCGATCCTGGGCGGCGCC GTCGGCGGCATCGCCTACCTGGGCATGGTCGGCTACCACCACCCCGGCCGCTCGTTCGGACCGTACCC GGAGTACCGGTAC	
Protein (299 aa)	
MSADAYNDKHGDEYINLEAHDGLSTDVGSLNGVTHVPYQVKSAFARQMMAEFFATFICVVVGLACTAQ VTLSSGTAGSFVTVALAWGFAYFLGITVGGGVSGANLNPTVTVALALLGMLPWKKVPFYILAQTVASY VAAMVVYILYRPMLNIVDPERTSTNAIFATYPHENVGNFTCFLTEFFAAALLVVGFLALLDEHNRPIG RAAVPAAIGVLVSAIAMGFGMNTGLAINAARDLGPRLLMLTVGFGSRVFSLSHYYFWVPLVAPILGGA VGGIAYLGMVGYHHPGRSFGPYPEYRY	

estExt_fgenesh1_pg.C_860075 (PsAq27)	Protein ID 140537
DNA sequence	
ATGGGGCCTCGACTCAGCGTCGAACCTACTGTGGACTGCAGCGCTGCGAGTGCGGCAGCTCCTTACTC CACCTTCGCCGCGACACCGTACAGCTGGCGTGGATTCATCGACCGCTGCGAGCGCTCGCTTCATCGCA CCAACCGCAGTCAGTACACCCGCGACGAGAGCCATCACGAGAAATGGTTTGTGACGAACCGCGTGCAC GTCCGCGAGTGCCTCACCGAGTTCCTCGGCACCTTCGTCATGATCTGCTTCGGCGTGGGGGTCAACAA CCAAGTCGAGCTCTCGGACGATTCCAACGGCACGTGGCTCGGTATCAACATAGCGTGGGGCATCGCTG TGCTCATGGGCGTCTACTGCTCCGAAGGTGTCAGCGGAGCGCACTTGAACTCCGCCGTCACGTTCGCC CACGCCTTCTACGGCCGTCTCGCGTGGTGGAAGCTGCCAGGTTACTGGATCTCGCAGGTGTTGGGCGC GTTCGTGGGCGCTGCCGCCATCTACCTGCTCAACTACCAGAAGATCCTGAAAACCGACCAGAACAAGG AGACCACCAAGTCCAACTTTGCGACGTATCCAAGCTCCGACATCAACAACCTCACGGCCTTCTACACG GAGGCACTCGCAACGGGCATGCTGCTGCTGTGCATCTACGCCATCACGGACCAGAACAACCGCTCGCC GGGGACGGTGGGTACGCCGTTCGCCTTTGCACTGATGATCATGGCGCTGGGTATGAGCTTCGGCATGA ACACGGGGTATGCGATGAACCCTGCGCGCGACTTCGGCCCTCGACTGCTCACTTACGCGGTCGGCTAC GGCTCGAAGGTCTTCACGACGCGCAGCTACTACTTCTGGATCCCGATCTGCGGTCCGCTTGTTGGAGG CGTCATTGGTACGCGTTCTTGGTGCAAGTGCAGCACCCGCACGAGCACG	
Protein (316 aa)	
MGPRLSVEPTVDCSAASAAAPYSTFAATPYSWRGFIDRCERSLHRTNRSQYTRDESHHEKWFVTNRVH VRECLTEFLGTFVMICFGVGVNNQVELSDDSNGTWLGINIAWGIAVLMGVYCSEGVSGAHLNSAVTFA HAFYGRLAWWKLPGYWISQVLGAFVGAAAIYLLNYQKILKTDQNKETTKSNFATYPSSDINNLTAFYT EALATGMLLLCIYAITDQNNRSPGTVGTPFAFALMIMALGMSFGMNTGYAMNPARDFGPRLLTYAVGY GSKVFTTRSYYFWIPICGPLVGGVIRRHWYAFLVQVQHPHEHVE	

estExt_fgenesh1_pg.C_950075 (PsAq29)	Protein ID 141227
DNA sequence	
ATGTCTGAGTACGGATCTCTCAAGCCTGGATACGCTGACCTCGAGGGCAACCCGCGACTTCCGTTCTA CGCCATCGCGAACCCCGATGTGCGCGCCTACCTCGCCGAGTTCGTGGGCACCTTCATCCTGGTTCTCA TCGGTGACGGCTCCGTGGCCCAGTACGTGCTTGGCGGTGGTGACGCTGGCCACTACCTGTCTGTCAAC CTGGCCTGGGGCATCGCTTTGCTGTTCGGCATCCACTTCTCCGGCGGCGTCAGCGGTGGCCACCTTAA CCCCGCGGTCACGCTGACCCTCGCTACCTTCGGCCGCTTCGAATGGTACAAGCTGCCCGGTTACTTCA TCGCGCAGACACTCGGTGCCTTCGCGGCCGCCTGGGTGGTCTTCATCGTGTACTACCCGTGGTTCGAC GTCCAGGACCCGGAGCGTGCCACCACCCAGGGCATCTTCGCCACGTACCCGAACGAGCAGATCCCCAA CTGGTCCGGCCTGGTCAACGAGATTGTCGGCACTGCCCTGCTCCTCAGCGGCATCTTTGCTGTCGGTG ACCAGCTCAACAAGCCGGCCAGCCCGTACACGTTCCCGGCCGCTGTCGCTCTCATGCTGACCTGCGTG GGCATGGCCTTCGGCCTCGACACGGGCTACGCGCTCAACCCGGCCCGCGACTTCGGCCCTCGCCTCTT CACCTTCTTCGCTGGCTGGGGCTGGAAGGTGTTCACGGCCCGCAGCTTCTACTTCTGGATCCCCATCA TCGGCCCCTTCATCGGTGGTCTGCTTGGTGCTGGCCTGTACGTCGGTCTCATTGAGAACTTCCACCCG CGCGAG	
Protein (274 aa)	
MSEYGSLKPGYADLEGNPRLPFYAIANPDVRAYLAEFVGTFILVLIGDGSVAQYVLGGGDAGHYLSVN LAWGIALLFGIHFSGGVSGGHLNPAVTLTLATFGRFEWYKLPGYFIAQTLGAFAAAWVVFIVYYPWFD VQDPERATTQGIFATYPNEQIPNWSGLVNEIVGTALLLSGIFAVGDQLNKPASPYTFPAAVALMLTCV GMAFGLDTGYALNPARDFGPRLFTFFAGWGWKVFTARSFYFWIPIIGPFIGGLLGAGLYVGLIENFHP RE	

estExt_fgenesh1_pg.C_860082 (PsAq30)	Protein ID 140544
DNA sequence	
ATGGTCAAGTACGCGACGGAGGACGTGGCCCCGACGCCCCAACAAGAGACTCTCGTGATCAACGCTAA AGCAGACAAGGTCGAGAGCGGCTACGCCAACAGGCCCACGACCCCCAAGGCGACGGACCTGGTGCCGA TCATGGCCGAGAAGCGCACGGGCATGTACCGGCTCGCAGTGCAGAGCGTGCATCTGCGCGAGTGCCTG GCCGAGTTCCTCGGCACGTACGTCATGCTGGTCTTCGGCATGGGCGTCAACAACCAAGTGACCAACTC GCAGGACGCCAACGGCACGTGGCTCAGCATCAACATTTGCTGGGGCATCGGCGTCATGATCGGCGTCT ACTGCTCCGAGGGCATCAGCGGCGCCAACTTGAACACTGCTGTGACGTTCGCCCACTGTGTCTACGGC CGTCTACCGTGGTGGAAGGCCCCCGGTTACATGTTCTCGCAGTCCCTGGGCGCCTTCTGCGGCGCGTT CACGATCTGGGTCATGCAGTGGCAGAACTTGAACAAGATCGACCCCAACCGTGAGACCACGCAGAGCA GCTTCGCCACGTACCCTGCTGACGGCATCTCCAACTACACGGCCTTCTACACGGAGTTCGTGGGCACG GCCATGCTGGTGCTTGCCATCTACGCCATCACCGACCAGAAGAACCGTCCCGCCGGCAAGTACGGCAA CGCCTTCGCCTTCGCTCTGATGGTCATGGCTCTGGGTATGGCCTTCGGCATGAACACTGGCTACGCTG TGAACCCGGCGCGTGACTTCGGCCCTCGTCTCTTCACGTTGTGCGCTGGCTGGGGAACGAAGACGTTC ACTTCGCATCATTACTACTTCTGGATCCCCATTGTGTCGGACCTGCTTGGCGGCGTAGCTGGTGCCGG TCTCTATCGTCTCATGGTGGAGATCCACCACCCCCAGACGCCTCTGCTC	
Protein (311 aa)	
MVKYATEDVAPTPQQETLVINAKADKVESGYANRPTTPKATDLVPIMAEKRTGMYRLAVQSVHLRECL AEFLGTYVMLVFGMGVNNQVTNSQDANGTWLSINICWGIGVMIGVYCSEGISGANLNTAVTFAHCVYG RLPWWKAPGYMFSQSLGAFCGAFTIWVMQWQNLNKIDPNRETTQSSFATYPADGISNYTAFYTEFVGT AMLVLAIYAITDQKNRPAGKYGNAFAFALMVMALGMAFGMNTGYAVNPARDFGPRLFTLCAGWGTKTF TSHHYYFWIPIVSDLLGGVAGAGLYRLMVEIHHPQTPLL	

estExt_fgenesh1_pg.C_950074 (PsAq32)	Protein ID 141226

DNA sequence

```
ATGACCGAGTACGGCACCAGCAAGCACCGCTACGCCGACCTCGAGAGCAACTCGCGCCTGCCGTTCTA
CCACGTGCACAACCCCAACTTCCGCGCCTACCTGGCCGAGTTCGTTGGCACCTTCATCCTCGTGCTCA
TCGGCGATGGCTCCGTGGCCCAGTACGTTCTGGGCGGGGGCGGTGCCGGCCACTACCTGTCGGTGAAC
CTGTGCTGGGGCATCGCGCTGCTCTTCGGCATTCACTTCTCCGGCGGTGTGAGCGGTGGCCACCTCAA
CCCGGCCGTGACCCTCACGCTGGCCCTGTTCAAGCGTTTCGAGTGGCACAAGGTGCCCGGCTACTTCG
TGGCCCAGACCGCCGGCGCCTTTGCCGCCGCCATCCTCGTCTTCATCGTGTACTACCCGTGGTTCGAC
ATCCACGACCCGGAGCGCGTGACCACGCAGGGCATCTTCGCTACCTACCCGAACGAGAAGATCCCCAA
CTGGTCAGCTCTGGCTAACGAGATCATCGGCACGGCCCTGCTGGTCAGCGGCATCTTCGCGGTCGGCG
ACCAGCTCAACAAGCCGGCCAGCCCCTACAGCTTCCCGGGTGCCGTGGCCCTGATGCTCACGTGCATT
GGCATGGCCTTCGGCCTCGACACGGGCTACGCGCTCAACCCGGCCCGCGACTTCGGCCCTCGCCTCTT
TACCTTCTTCGCTGGCTGGGGCTGGCAGGTGTGGACGCTGCGCGGATGCTACTTCTGGATCCCCATTG
TTGGCCCGTTCGTCGGCGGCATCCTTGGTGCTGCCACCTACGTGGGCCTCATCGAGGCGCACCACCCG
CCGCAG
```

Protein (274 aa)

```
MTEYGTSKHRYADLESNSRLPFYHVHNPNFRAYLAEFVGTFILVLIGDGSVAQYVLGGGGAGHYLSVN
LCWGIALLFGIHFSGGVSGGHLNPAVTLTLALFKRFEWHKVPGYFVAQTAGAFAAAILVFIVYYPWFD
IHDPERVTTQGIFATYPNEKIPNWSALANEIIGTALLVSGIFAVGDQLNKPASPYSFPGAVALMLTCI
GMAFGLDTGYALNPARDFGPRLFTFFAGWGWQVWTLRGCYFWIPIVGPFVGGILGAATYVGLIEAHHP
PQ
```

G.6 JGI DB analysis – *P. ramorum*

fgenesh1_pm.C_scaffold_114000003 (PrAq2)	Protein ID 72234

DNA sequence

```
ATGATCCGCGACAGCATGGACAAGATGGAGGCGATGCTTCGAGGCAGCATGAGCAAGCTGGACAGCAA
CTACCTGGGCGAGCCCGATGTGGCTATGATGTCGACGGAGCATCTACCCGCAGCAGCGTATCAAGCCA
AGTCACCGCTCATCCGTGAGTGTATGGCCGAGTTCCTCGGCATGTTGGTCATGATGTCGTTTGGCACC
GGCGTCGTCGCCCAAGTCGTGCTAAGCGAAGGCACGAAAGGAGAATTCATCAGCATCAACTTGTGCTG
GGGGCTCGCGGTTCTCTTCGGAATCCACGTCTGCGGTGGCGTCTCCGGTGGCCACCTGAACCCCGCGG
TGACTGTGACGTTGGCGCTGTTCCGCCGTTTTGAGTGGAAGAAGGTGCCGTTCTACGTGCTAGCTCAA
ATGCTAGGAGCTTTTGCTGCCTCGGCCATTGTCTGCATGGTGTACGCCCCAATGATCAACGTCGTCGA
CCCGCACAAGACTACAACTCAGGGGATATTTGCAACGTATCCGTACAGCGACGACGTCCCTATCGGGA
CTTGCTTCCTAACCGAAGTCGTTGGTACCGCACTGCTGGTCGGCTGTTTGTTCGCCATCGGAGACGAA
ATGAACAAACCTGCCAGCCCCTTCACGCAGCCAGCAGCGGTGGCGCTTTTGGTGGTGGCGATCGGGAT
GGCTTTCGGCATGAACACCGGGTACGCTATCAACCCGGCTCGCGACCTCGGCCCTCGTCTCTTCTCGC
TCTGTGCTGGCTGGGGGTCCAAGGTGTTCACGCTGCGTGACTCGTACTTCTGGGTGCCGATTGTCGCC
CCGTTGCTTGGTGGCCCGATTGGTGCTGGCATGTACGTCGTCATGGTGGAGCAGCACCACCCGCGCGC
G
```

Protein (295)

```
MIRDSMDKMEAMLRGSMSKLDSNYLGEPDVAMMSTEHLPAAAYQAKSPLIRECMAEFLGMLVMMSFGT
GVVAQVVLSEGTKGEFISINLCWGLAVLFGIHVCGGVSGGHLNPAVTVTLALFRRFEWKKVPFYVLAQ
MLGAFAASAIVCMVYAPMINVVDPHKTTTQGIFATYPYSDDVPIGTCFLTEVVGTALLVGCLFAIGDE
MNKPASPFTQPAAVALLVVAIGMAFGMNTGYAINPARDLGPRLFSLCAGWGSKVFTLRDSYFWVPIVA
PLLGGPIGAGMYVVMVEQHHPRA
```

fgenesh1_pg.C_scaffold_46000075 (PrAq4)	Protein ID 80164

DNA sequence

```
ATGCGGGTGCACCTGAACGCGGAGCCCAGCGCCGGCCACATGACCCTCGAGGGCGTCGCAACGCCCGG
TTACTCGGCCGTCCCCCCCACGCCCTTCACATGGCGCGGGTTCATCGACCGCTGCGAGCGCGCGCTGC
ACCACACCAACCGCAGCCAATACACGCGCGACGGAGACCACCACAAGAAGTGGTTCGTCACCAACAGC
AGCCACATCCGCGAGTGCCTCGCCGAGTTCCTAGGGACGTTCGTCATGGTCTGCTTCGGTATGGGCGT
CAACAACCAAGTGTCGCTGTCCGAGGAGGCCAACGGCACGTGGCTGAGCATCAACATGGCGTGGGGCA
TCGCTGTGCTCATGGGCGTCTACTGCTCGGAAGGAGTGAGCGGCGCGCACCTCAACTGCGCCGTCACG
CTCGCCCACGCCGTGTATGGTCGTCTGCCGTGGTGGAAGCTCCCGGGTTATTGGCTGTCGCAGGTGGT
GGGCGCGTTCATCGGCGCCGCCTCCATCTACGTGCTCGACTACCAGAAGATCCACAAATTGGACCCGG
AGCTGGACACCACCCAAAGCAACTTCGCCACGTACCCGAGCGACGACATCAGCAACCTCACGGCCTTT
TACACGGAAGCCCTGGCAACGGGGATGTTGTTGCTGTGCATCTACGCCATCACGGACCAGAACAACCG
ATCGCCCGGTACCGTCGGCACGCCGTTCGCCTTCGCGTTGATGATCATGGCGCTCGGTATGAGCTTTG
GTATGAACACAGGGTATGCTATGAACCCGGCACGTGACTTTGGTCCTCGATTGTTAACGTACGTGGTG
GGCTATGGCTCCAAGGTGTGGTCGACGCGCAATTACTACTTCTGGATCCCCATCTTCGGCCCTCTGGT
GGGCGGGGTTGTCGGCGCAGGACTGTACACGTTCCTCGTGCAGGCGCAGCACCCGCACGAGCACGAGT
CT
```

Protein (318 aa)

```
MRVHLNAEPSAGHMTLEGVATPGYSAVPPTPFTWRGFIDRCERALHHTNRSQYTRDGDHHKKWFVTNS
SHIRECLAEFLGTFVMVCFGMGVNNQVSLSEEANGTWLSINMAWGIAVLMGVYCSEGVSGAHLNCAVT
LAHAVYGRLPWWKLPGYWLSQVVGAFIGAASIYVLDYQKIHKLDPELDTTQSNFATYPSDDISNLTAF
YTEALATGMLLLCIYAITDQNNRSPGTVGTPFAFALMIMALGMSFGMNTGYAMNPARDFGPRLLTYVV
GYGSKVWSTRNYYFWIPIFGPLVGGVVGAGLYTFLVQAQHPHEHES
```

fgenesh1_pm.C_scaffold_716000001 (PrAq5)	Protein ID 72425

DNA sequence

ATGACTTCCAACCAAGACACAGCGGCTATTTCGTACGACGAGCTGATCGATAAGGTCGAGACGGGCTA
CGTCGACGCGCCACTGACGCCCAAGGCCGGCTCCGCGTCGCAGGGCAAGGCCGTCTCGCGCTACGCTG
TGCAGAGCGTGCACATGCGCGAGTGCCTGGCCGAGTTCCTCGGCACCTTCATCATGATCGTCTTCGGC
ATGGGCGTGAACAACCAAGTGGTGAACTCCGAGGACAAGAACGGCACGTGGCTCAGCATCAACATGTG
CTGGGGCGTGGCCGTACTCATCGGCGTGTACTGCTCCGAGGGCGTCAGTGGCGCCAACCTCAACACGG
CCGTGACCCTGGCCCACTGCGTGTACGGCCGCCTGCCGTGGTGGAAGGCGCCGGGCTACATGGTGTCG
CAGGTGCTGGGCGCCTTCTGTGGCGCCTTCGTCATCTACCTCATGCAGTACCAGAACCTGAACGTGAT
CGACCCGAACCGCGAGACGACGCAGAGCAGCTGGTCCACGTACCCTAGCGACCAGATCTCCAACTACA
CGGCATTCTACACGGAGTTCATCGGCACGGCCATGCTGGTGCTGAGCATCTACGCCATCACGGACCAG
CGCAACCGGCCGGCCGGTCCCGTGGGCGCCGCGTTCGCCTTCTGTCTCATGATCATGGCTCTGGGTAT
GGCGTTCGGCATGAACACGGGATACGCCGTGAACCCTGCTCGTGACTTCGGCCCTCGCCTGTTCACGT
TCTGCGCCGGCTGGGGCTCCAAGGTGTTCACGACGCGCAACTACTACTTCTGGATCCCGCTCGTGGCC
GACTCGCTGGGCGGCGTGTGCGGCGCGGGTCTGTACCGCCTGCTCGTGGAGATCCACCACCCTCCCCT
CCCGCACCAGAAC

Protein (299 aa)

MTSNQDTAAISYDELIDKVETGYVDAPLTPKAGSASQGKAVSRYAVQSVHMRECLAEFLGTFIMIVFG
MGVNNQVVNSEDKNGTWLSINMCWGVAVLIGVYCSEGVSGANLNTAVTLAHCVYGRLPWWKAPGYMVS
QVLGAFCGAFVIYLMQYQNLNVIDPNRETTQSSWSTYPSDQISNYTAFYTEFIGTAMLVLSIYAITDQ
RNRPAGPVGAAFAFCLMIMALGMAFGMNTGYAVNPARDFGPRLFTFCAGWGSKVFTTRNYYFWIPLVA
DSLGGVCGAGLYRLLVEIHHPPLPHQN

fgenesh1_pm.C_scaffold_102000001 (PrAq6)	Protein ID 72182

DNA sequence

ATGTCACCCGGCGCCTACCAGAAGAACGACGGAGACTACGTCGAGTTGGAAGCCAACGACGGACTCTC
CACAGTTGTGAGTTCTCACGCCCAGGTGGAGGACGTGTCGTACGCCCCGTACCAAGTCAAGAGCCCAT
TCGCCAAGCAGTTGATGGCCGAATTCCTCGCCACCTTCGTCTGCGTGGTCTTCGGACTGGCTTGCACG
GCGCAAGTCACCCTGAGCGAGGGCAGTGCCGGCAGCTTCGTGACCGTCGCTCTGGCGTGGGGCTTTGC
CTACTTCCTCGGCATCTCGGTCGGCGGTGGCGTCTCCGGCGCCCACCTCAACCCGACCATCACTGTGG
CGCTCGCCGTCATTCAAATGTTCCCGTGGAAGAAGGTCCCGATCTACATCCTCGTGCAGACCGTTGCC
GCCTACGTGTCGGCTCTCGTGGTCTACATCCTCTATCGCCCGTTGCTCAACGAGGTGGACCCCGACCG
CACCACGAGCCACGCCATTTTCGCCACGTACCCGCACGCGAACGTGGGCAACTTCACGTGCTTCCTGA
CCGAGTTCTTTGCTGCTGCGGTGCTGATCCTCGGTCTCTTGGCGCTCCTGGATCAGCACAATCGCCCC
ATCGGCCGCCGCGCCGTTCCGGCCGCTGTCGGTGTTCTCGTGAGCGCGATCGCCATGGGTTTCGCCGT
GAACACGGGTCTGGCCATCAACGCCGCTCGCGACCTCGGGCCTCGCCTCCTCATGCTCACCATGGGCT
GGGGCTCACACGTCTTCTCGTTGAACCACTACTACTTCTGGGTCCCGCTCGTGGCCCCCATTCTGGGT
GGCGCCGTCGGCGGCATCGCGTACGAAGGCATGGTCGGCTACCACCACGTCGGCCGGTCTCGCGGACC
GTACCCGGAGTATCGTTTC

Protein (301 aa)

MSPGAYQKNDGDYVELEANDGLSTVVSSHAQVEDVSYAPYQVKSPFAKQLMAEFLATFVCVVFGLACT
AQVTLSEGSAGSFVTVALAWGFAYFLGISVGGGVSGAHLNPTITVALAVIQMFPWKKVPIYILVQTVA
AYVSALVVYILYRPLLNEVDPDRTTSHAIFATYPHANVGNFTCFLTEFFAAAVLILGLLALLDQHNRP
IGRRAVPAAVGVLVSAIAMGFAVNTGLAINAARDLGPRLLMLTMGWGSHVFSLNHYYFWVPLVAPILG
GAVGGIAYEGMVGYHHVGRSRGPYPEYRF

fgenesh1_pm.C_scaffold_46000011 (PrAq7)	Protein ID 71777
DNA sequence	
ATGGTCAAGTACGCGACCGAGGACGTGGCCCCAACACCCAAGCAGGACACTCTCGTGATCAACGCCGA AGCCGACAAGATCGAGAACGGATACGCCAACATGCCCAACACCCCCAAGGCGACGGATTTTATGCCGA TCCTGGCCGAGAAGCGCACGGGCATGTACCGGTTCGCCGTGCAGAGCGTGCACTTGCGCGAGTGCTTC GCCGAGTTCCTCGGTACCTTCGTCATGATCGTCTTCGGCATGGGCGTGAACAACCAAGTGACCAACTC GGAGAACGCCAACGGCACGTGGCTCAGCATCAACATTTGCTGGGGCATCGGTGTCATGCTTGGCGTCT ACTGCTCGGAAGGTATCAGCGGCGCCAACCTCAACACGGCCGTGACATTGGCTCACTGCGTGTACGGT CGACTACCGTGGTGGAAGGCGCCTGGCTACATGTTCTCACAGTTCCTTGGCGCCTTCTGTGGCGCCTT CACGATTTGGGTGATGCAGTGGCAGAACTTGAACAAGATCGACCCGAACCGCCAGACGACGCAGAGCA GCTTCTCCACGTACCCAGCTGAGGGTATCTCGAACCTGACTGCCTTCTACACGGAGTTCATTGGTACA GCCATGCTGGTGCTTGCCATCTACGCCATCACGGACACCAAGAACCGCCCCGCCGGCAAGTACGGCAA CGCCTTCGCCTTCGCCTTGATGATCATGGCTCTGGGTATGGCGTTCGGCATGAACACGGGGTACGCCG TGAACCCCGCGCGTGACTTCGGCCCTCGTCTGTTCACGTTCTGCGCCGGCTGGGGCTCCAAGGTCTTC ACATCGCACCACTACTACTTCTGGATCCCCATTGTCTCGGACCTGCTAGGAGGCGTTTGTGGTGCTGG GCTCTATCGTCTTCTCGTGGAGATCCACCATCCCCAGACGCCTTTGCTT	
Protein (311 aa)	
MVKYATEDVAPTPKQDTLVINAEADKIENGYANMPNTPKATDFMPILAEKRTGMYRFAVQSVHLRECF AEFLGTFVMIVFGMGVNNQVTNSENANGTWLSINICWGIGVMLGVYCSEGISGANLNTAVTLAHCVYG RLPWWKAPGYMFSQFLGAFCGAFTIWVMQWQNLNKIDPNRQTTQSSFSTYPAEGISNLTAFYTEFIGT AMLVLAIYAITDTKNRPAGKYGNAFAFALMIMALGMAFGMNTGYAVNPARDFGPRLFTFCAGWGSKVF TSHHYYFWIPIVSDLLGGVCGAGLYRLLVEIHHPQTPLL	

fgenesh1_pg.C_scaffold_114000013 (PrAq8)	Protein ID 84474
DNA sequence	
ATGCCGAGCAACGACGACAAGATCGAGAGCGGCTACGTCGGCGTGCAAGAACCAGATATGCTGATGGA CCTCCAGCCTGCCAAGCCGTTCCAAGTTAAGTCCCCGCTCTTCCGCGAATGTCTGGCCGAGTTTCTCG GCACCATGGTCATGATCATGTTCGGCGACGGCGTCGTGGCCCAGGTCGTGCTCAGCGAGGGCACCAAG GGCGAGTACATCAACATCAACTTGTGCTGGGGTCTCGGTGTGCTCTTCGGCATTCACGTCTCTGGTGG AGTCTCCGGTGCTCACCTGAACCCCGCGGTGACCACGACGCTCGCTTGGTTCGGACGCATGGAGTGGC GCAAGGTCCCGTACTACGTGATCTCGCAGATCCTCGGCGCCTTCGTGGCTGCATTCATCGTCTGGGTC GTGTACTACCCGATGTTCAATGTGATCGACCCCGAGAAGACCTCGACCCAAGGTGTGTTCGCGACGTA CCCGTTCAGTAACGACGTCCCCGTCGGAACCTGCTTCCTGACAGAAGTGGTGGGCACGGCACTGCTCA TGGGCTGCATCTTTGCTATCGGCGACGAGCTCAACAAGCCCGCTAACCCGTACACGCAGCCCGGTGCA GTGGCATTGCTGGTGGTGGCTATTGGTATGGCGTTTGGTATGAACTCCGGCTACGCTATCAACCCTGC GCGTGACTTCGGCCCGCGTTTCTTCTCCTGGATGGCCGGCTGGGGGTCGGAGGTCTTCACGCTGCGCG ACTCGTACTTCTGGGTGCCCATTGTCGGCCCGCTTCTCGGCGGACCTATTGGTGCTGGCATCTACGTT GGCTTGATCGAGCACCACCACCCGCGCGAGTACAAGCACCCGCTTGAGGGT	
Protein (289 aa)	
MPSNDDKIESGYVGVQEPDMLMDLQPAKPFQVKSPLFRECLAEFLGTMVMIMFGDGVVAQVVLSEGTK GEYININLCWGLGVLFGIHVSGGVSGAHLNPAVTTTLAWFGRMEWRKVPYYVISQILGAFVAAFIVWV VYYPMFNVIDPEKTSTQGVFATYPFSNDVPVGTCFLTEVVGTALLMGCIFAIGDELNKPANPYTQPGA VALLVVAIGMAFGMNSGYAINPARDFGPRFFSWMAGWGSEVFTLRDSYFWVPIVGPLLGGPIGAGIYV GLIEHHHPREYKHPLEG	

fgenesh1_pm.C_scaffold_114000004 (PrAq9)	Protein ID 72235
DNA sequence	
ATGGCCGAGTACAAGCCTGGATACGGAGGTTATGCCGATCTCGAAGGAAATGGCCGTTTGGCGCCCTA CGCCATCCGCAACCCGGACATGCGCGCGTACATGGCCGAGTTCGTCGGCACCTTTATCCTCGTGCTGA TCGGAGACGGATCCGTGGCTCAGTTCGTGCTCAGCAAAAGAGCTGCTGGCGACTACCTCTCGGTGAAC CTGTGCTGGGGTATCGCGCTGCTGTTTGGCGTTCACTTTTCGGGTGGAGTGAGCGGTGGCCACCTGAA CCCTGCAGTGAGTGTGACTATGGCTCTCTTCAAGCGCTTCGAGTGGCGTCGGGTTCCGGGTTACATCA TCGCGCAGACGCTGGGCGCCTTCGTTGCTGCTTTCGTGCTCTACATCGTGTACTACCCGTGGTTCGAC ATTGTCGACCCGGAGCGTGAGTTCACGCACGGCATCTTCGCGACGTACCCGAACCCACAGATCCCCAA CTGGGCTGCGTTTGCCAACGAAGTCGTCGGTACAGCTCTGCTGGTTGGCGGCATCTTCGCGCTGTGTG ACCAGATCAACAAGCCAGCCAGTCCCTACAGCTTCCCGGGTGCGGTTGGCCTTCTCCTCACGGGCATT GGCATGTCGTTCGGTCTCAACACGGGGTATGCGCTCAACCCTGCTCGTGACTTCGGCCCACGTCTTCT GACGTTCTTCGGAGGCTGGGGCTGGAAGGTGTTCTCGTCGCACGGCGGGTACTTCTGGATCCCCATTG TCGGCCCGTTCATGGGCGCTGTGCTGGGCGGCGCCATGTACGTGGGCCTCGTCGAACTGCACCACCCA CCCCAG	
Protein (274 aa)	
MAEYKPGYGGYADLEGNGRLAPYAIRNPDMRAYMAEFVGTFILVLIGDGSVAQFVLSKRAAGDYLSVN LCWGIALLFGVHFSGGVSGGHLNPAVSVTMALFKRFEWRRVPGYIIAQTLGAFVAAFVLYIVYYPWFD IVDPEREFTHGIFATYPNPQIPNWAAFANEVVGTALLVGGIFALCDQINKPASPYSFPGAVGLLLTGI GMSFGLNTGYALNPARDFGPRLLTFFGGWGWKVFSSHGGYFWIPIVGPFMGAVLGGAMYVGLVELHHP PQ	

fgenesh1_pm.C_scaffold_114000002 (PrAq11)	Protein ID 72233
DNA sequence	
ATGGCGACCCCTTCACACGCAGCGGACAAGATCGAGCACGGCTACATGGACCTGGAAGCCCACGAGGC CGGGGTGAACGGCGGCCCCGTCGTGAACCCGAGTCGGTTCGCGCTCAAGAACCCGCACGTCCGCGAAG TCTTCGCCGAGTTCGTGGCCGTCTTCGTCATGATGAGCTTCGGTCTGGGCGCGACGTGCCAGTCCATC CTGAGCGACAGCAAGTACGGCGACTTCACGCAGATCGTCTTCGGCTGGGGCATCGGCGTGCTCTTCGG CATCCACATCGCCGGCGGCGTCTCCGGTTCCCACTTGAACCCGGCCATCACCACCACCGCCGCACTCT TCGGCATGTTCCCGTGGAAGAAGGTGCCTTCCTACGTGGCCGCACAGACGCTGGCCTCCTACTTAGCC GCTCTTTTCATCTACGTGCTGTACCGGCCGCTGCTCAATGTGGCCGACCCGGAACGGCAGACAACTCA CGCCATCTTCGCCACGTACCCGAACGAGTACGTGTCCAACTTCATCGCGTTCCTGACCGAGATGTTCG CGACGGCGCTGCTGGTGGGCGGCATCTTCGCCGTGCTGGACCAGCACAACCGACCGGCCAGTCCGTTC AGCGCCCCCAGCGCCGTGGCGCTGCTGGTGGTGGGCATCGGCATGGCCTTCAGCGTCAACACGGGCTG TGCCATCAACCCGGCGCGCGACTTCGGCCCGCGCCTCATGATGCTCACCGCCGGCTGGGGCTCGCACG TCTTCTCGCTCAACCACTACTACTTTTGGGTGCCCATCGTGGGCCCCACCGTCGGCGGCGCGCTCGGT GGCCTCGTGTACGTGGCGCTCGTGGGCCACCACCACCCGCACAAGGCGGTAGTCGTTCGCAAGCACGG CGAGCAGTTCCACTTC	
Protein (300 aa)	
MATPSHAADKIEHGYMDLEAHEAGVNGGPVVNPSRFALKNPHVREVFAEFVAVFVMMSFGLGATCQSI LSDSKYGDFTQIVFGWGIGVLFGIHIAGGVSGSHLNPAITTTAALFGMFPWKKVPSYVAAQTLASYLA ALFIYVLYRPLLNVADPERQTTHAIFATYPNEYVSNFIAFLTEMFATALLVGGIFAVLDQHNRPASPF SAPSAVALLVVGIGMAFSVNTGCAINPARDFGPRLMMLTAGWGSHVFSLNHYYFWVPIVGPTVGGALG GLVYVALVGHHHPHKAVVVRKHGEQFHF	

fgenesh1_pg.C_scaffold_114000015 (PrAq12)	Protein ID 84476
DNA sequence	
ATGGACCCGCTCTCATACAGCTACAAGATCGACGAAGATTACACCGCCGTTCAATCTCCTCGCCAGAT TGTGCTTGACGCTCCAGCTATTCGCCCCTACCAATTCAAGTCCCCACTACTACGAGAGTGCCTCGCCG AGTTCCTAGGAACCCTCATACTGGTCCTCTTCGGAGACGGAGCCGTCGCTCAAGTCACGCTCAGTGAA AACACCAAGGGAAACTACACCACGCTCTGCCTCGGATGGGGCGTTGGCATCCTCTTCGGCATCCACGT ATCTGGAGGCGTCTCCGGCGCACACATCAACCCGGCCATCACCACCACGCTCGCGCTCTTCCGCCGCT TCGAGTGGCGCAAGGTCGTCCCGTACATCGTGGCGCAGACGCTGGGCGCCTTCGTGGCCGCCTTCCTC ATCTGGGCCGTCTACAAGCCGCTCTTCGACGCCGTCGACCCCGACAAGACTTCCACGCACGGTGTCTT CGCCACGTACCCCTACAGCAGCGACGTGTCCGTGGGGACATGCTTCTTGACGGAGGTGGTGGGCACGG CGCTACTGCTGTGCGGCCTCTTCGCTATCGGCGACGAGCTCAACAAGCCTGCCAACCCGTACTCTCAG CCAAGTGCCGTAGCGCTGTTGGTGGTGGGTGTCGGCATGGCGTTCGGCATGAACACGGGTTTCGCGTT GAACCCTGCGCGCGACTTCGGTCCTCGGCTGTTCACGCTGTGCGCTGGTTGGGGTTCACAGGTATTCA CGCTTCGCGATGCGTACTTTTGGGTTCCGATCGCGGCTCCGTTGTTGGGCGGTGCGATCGGTGCGGGG GTGTACGTCGGTTTGGTGGAGCACCATCACCCTCGAGAGTACAACCCGACACTCCCGAGCGCTGGCAG CTTCGCTGTGGCCAGCCACTCCTCGTACAAG	
Protein (305 aa)	
MDPLSYSYKIDEDYTAVQSPRQIVLDAPAIRPYQFKSPLLRECLAEFLGTLILVLFGDGAVAQVTLSE NTKGNYTTLCLGWGVGILFGIHVSGGVSGAHINPAITTTLALFRRFEWRKVVPYIVAQTLGAFVAAFL IWAVYKPLFDAVDPDKTSTHGVFATYPYSSDVSVGTCFLTEVVGTALLLCGLFAIGDELNKPANPYSQ PSAVALLVVGVGMAFGMNTGFALNPARDFGPRLFTLCAGWGSQVFTLRDAYFWVPIAAPLLGGAIGAG VYVGLVEHHHPREYNPTLPSAGSFAVASHSSYK	

fgenesh1_pg.C_scaffold_46000085 (PrAq13)	Protein ID 80174
DNA sequence	
ATGGAGGAAGCAAGCTTCGTGCACAGTGGAGACATCTTCTCCCTCACCGACAAGTCCGGCTACGTGAT CCCGTACCACCAGGCCCCAAAAACTGGAGGTGGAAACGTGGGGCACCGCCGACGGATGCCGCGCTCCT ACTCGACGCACTCGGTGCTGCACGAAACCAACCGCAGCCAACTCTGGATCCCCGGAAGTCAGCGAGAG CACATTCCGTTCGTGACCAAGAGCGTGCACATGCGCGAGTGTCTGGCTGAATTCCTGGGCACGCTCGT GTTCCTCTGTTTCGGCATCGGTGTCAACAACCAAGTCAATCTCTCGGAAGACGCCAACGGAACTTGGC TCAGCGTCAACATTTGCTGGGGCATCGGCGTCCTCATCGGGGTATACGTCGCGGAGGGGATCAGTGGC GCGCACCTCAACACGGCCGTGACGTTCACCCACGCTGTGTTCGGGCGTCTTCCGTGGTGGAAAGTGCC TGGTTACGTTCTAGCTCAGACAATGGGAGCGTTCTGTGCGTCGGCTCTCGTTTTCGTGCTTCACTACC AGCGTCTGATGCTCGAAGACCCGGACAGCATGACCACACAAGGCAACTTCGCCACGTACCCGCGTGAC AGCATCTCCAACGTCACCGCCTTCTACTCGGAGACATTGGGGACCGCAGTCCTTCTCATGGCCATTTT CGCCATCACGGACGAGCGCAACCGTGGCGCGGGTCCCGTCGGTACTCCGTTCGCGTTCGCTCTGCTCT TCATGGGCCTCGGTATGGCGCTCGGCATGAACACTGGCTACGCGTTGAACCCGGCGCGCGACTTCGGC CCTCGCCTGTTCACCCTCCTGGCCGGCTACGGCCCCAAGGTCTTCTCCTCGCACTCGCACTACTTCTG GATCCCTATCGTCGGTCCTCTCGTGGGTGGCGTCATGGGCGCGGGGTCCTACTTCTTCATCGTGCAGC TGCAGCACAACGACGACGAGACCGAAAGCGTCAGCGACGTGCTTTCG	
Protein (333 aa)	
MEEASFVHSGDIFSLTDKSGYVIPYHQAPKTGGGNVGHRRRMPRSYSTHSVLHETNRSQLWIPGSQRE HIPFVTKSVHMRECLAEFLGTLVFLCFGIGVNNQVNLSEDANGTWLSVNICWGIGVLIGVYVAEGISG AHLNTAVTFTHAVFGRLPWWKVPGYVLAQTMGAFCASALVFVLHYQRLMLEDPDSMTTQGNFATYPRD SISNVTAFYSETLGTAVLLMAIFAITDERNRGAGPVGTPFAFALLFMGLGMALGMNTGYALNPARDFG PRLFTLLAGYGPKVFSSHSHYFWIPIVGPLVGGVMGAGSYFFIVQLQHNDDETESVSDVLS	

fgenesh1_pm.C_scaffold_102000002 (PrAq15)	Protein ID 72183

DNA sequence

ATGCCTTCGGCCACCTTCAACGAGAAGCGCGAGCTGGGATACGTCGACCTGGAAGCCCATGACGGCCA
CTCCACGACGAGGAGCTCTCATGCCCACGTGGACGACGTCGCCCTCGCTCCGTACCAAGTCAAGAGTC
AATTCGCCAAGGAAATGATGGCCGAGTTCCTCGCCACCTTCATCTGCATGCTCTTCGGCCTCTCCTGT
ATGGCCCAAGTCACCCTTAGCGGAGGCAGCGCCGGCAGCTTCGTGACCATCGCGTTGGCCTGGGGTTT
CGCCTACTTCCTCGGTATCTCAATTGGCGGCGGCGTGTCCGGTGCCCACCTCAACCCGACGGTGAGCG
TGACGCTCGCTCTGCTCAAGCTGCTGCCGTGGAAGAAGGTTCCGTTCTACATCCTCGTGCAGACCATT
GCTGCCTACGTGTCAGCTCTCGTGGTGTACATCCTGTATCGCCCCCTGTTTAACGAGGTGGACCCCGA
CCGCACCACGACGCACACCGTCTTCGCCACGTTCCCGCATGAAAACGTGGGCAACTTCACGTGCTTCC
TGACCGAGTTCCTCGCCGCGGCGCTGCTCATCCTCGGCATCCTGGCGCTCCTGGACCAGCACAACCGC
CCGATCGGCCGCCACGCGGTCCCTCCTGCCATTGGCGTGCTCGTGAGCGCGATCGCCATGGGTTTCGC
CATGAACACGGGTCTGGCTGCCAATGCTACCCGTGACCTCGGCCCTCGTCTCTTCATGCTCACCGCTG
GCTGGGGTACTCGGGTTTTCACGTTGAACCACTACTACTTCTGGATCCCGCTCGTGGCCCCGATCATG
GGCGGGGCTGCTGGTGCGTTCGTGTACGAGGCCATGGTCGGCTACCACCACGTCGGCCGATCTCGTGG
CCCGTACCCGGAGTACCGATTT

Protein (302 aa)

MPSATFNEKRELGYVDLEAHDGHSTTRSSHAHVDDVALAPYQVKSQFAKEMMAEFLATFICMLFGLSC
MAQVTLSGGSAGSFVTIALAWGFAYFLGISIGGGVSGAHLNPTVSVTLALLKLLPWKKVPFYILVQTI
AAYVSALVVYILYRPLFNEVDPDRTTTHTVFATFPHENVGNFTCFLTEFLAAALLILGILALLDQHNR
PIGRHAVPPAIGVLVSAIAMGFAMNTGLAANATRDLGPRLFMLTAGWGTRVFTLNHYYFWIPLVAPIM
GGAAGAFVYEAMVGYHHVGRSRGPYPEYRF

fgenesh1_pm.C_scaffold_46000012 (PrAq16)	Protein ID 71778

DNA sequence

ATGGCCCCATTCCCCTCCAACCAGACCCTGGACGTGGCGTCGGATAACCAAGTGCCGCAGCCTTTCGA
GAATCTCAATGATAAGCACGAAGCCAGCTATATCGACATGCCGCTGACTCCGAAGGCCAGCGGCGCGG
ACGCCAAACTCGCGGGCTTGTCCCGCTTCGCCGTGCAGAGCGTGCACTTGCGCGAGTGCTTCGCCGAG
TTCCTCGGTACCTTCGTCATGATCTTGTTCGGTATGGGCGTGAACAACCAAGTGACCAACTCGGAGAA
CGCCAACGGCACGTGGCTCAGCATCAACATGTGCTGGGGAATTGGCGTACTCATCGGCGTGTACTGTA
CGGAGGGTATCAGTGGCGCCCACTTGAACACGGCCGTGACCCTGGCCCACTGCGTGTACGGTCGCCTG
CCGTGGTGGAAGGCGCCGGGCTACATGATCTCGCAGGTGTTGGGGGCCTTCTGCGGAGCCTTCGTCAT
CTACGTCATGCAGTGGCAGAACCTGAACGTGATCGACCCGAACCGCGAGACGACGCAGAGCAGCTTCG
CCACGTACCCAAGTGACCAGATCTCCAACTACACGGCTTTCTACACTGAGGTTGTTGGTACAGCCATG
CTGCTACTCGGTATCTACGCCATCACGGACCAGCGTAACCGCCCGGCGGGTCCCGTGGGCGCTCCCTT
TGCCTTCTGCCTCTTGATCATGGCTTTGGGCATGTGTATCGGCATGAACACGGGCTACGCCATCAACC
CAGCGCGTGACTTCGGCCCCCGCCTGTTCACGAGCATCGCCGGCTGGGGCTCCAAGGTCTTCACCCTG
CGTGACCACTACTTCTGGATCCCGATCGTGGGCCCGCTGGTCGGTGGAGTGCTTGGCGCTGGGCTGTA
CACTGTCCTGGTGGAGATTCACCACCCGCCGCAGACGCTCCCT

Protein (309 aa)

MAPFPSNQTLDVASDNQVPQPFENLNDKHEASYIDMPLTPKASGADAKLAGLSRFAVQSVHLRECFAE
FLGTFVMILFGMGVNNQVTNSENANGTWLSINMCWGIGVLIGVYCTEGISGAHLNTAVTLAHCVYGRL
PWWKAPGYMISQVLGAFCGAFVIYVMQWQNLNVIDPNRETTQSSFATYPSDQISNYTAFYTEVVGTAM
LLLGIYAITDQRNRPAGPVGAPFAFCLLIMALGMCIGMNTGYAINPARDFGPRLFTSIAGWGSKVFTL
RDHYFWIPIVGPLVGGVLGAGLYTVLVEIHHPPQTLP

fgenesh1_pm.C_scaffold_46000010 (PrAq17)	Protein ID 71776

DNA sequence

ATGGAGTCGAGCTACAACAACACCCAACCCGTCATTGACGTCGCCGACAAGTCCAGCACCTACGCTCC
GTACGCCGACACGCCGTACTCGCCGCGTCAAGGCGACAGGTGCGAGCGCTCGCTGCACCACACCAACC
GCAGCCAGTATGACCGCGACGAGGACCACCACAAGCAGTGGTTTGTCACCAGGAACCCCCACCTGCGC
GAATGCTACTCGGAGTTCCTCGGTACTTTCGTCATGATCAGCTTTGGCATGGGCGTCAACAACCAGGT
CGTCCTCTCCGAGGATAAGGAGGGCACGTGGCTCAGCATCAACATGGCGTGGGGCATCGCTGTGCTCA
TGGGCGTCTACTGCTCGGAAGGAGTGAGCGGCGCGCACATGAACCCGTCCGTCACCTTGGCGCACGCT
GTGTACGGCCGTCTGGCTTGGCGCAAGGTACCCGGTTACATGCTTTCGCAGTTCCTCGGGGCTTTCGT
CGGTGCGTTCGCAATCTACCTGCTCGACTACCAGCGCCTTATGAAGGCTGACCCGGACAAGGAGACCA
TGTACCACAACTTCGCGACGCACCCCAACCCGGAGATCAACAACCTCACGGCCTTCTACACTGAGGCT
CTGGCTACGGGGATGCTGCTGCTGTGCGTCTACGCCATCACGGACCAGCACAACCGATCGCCCGGTAC
CGTCGGCACGCCGTTCGCCTTCGCTCTGATGATCATGGCGCTCGGTATGAGCTTCGGTATGAACACGG
GCTACGCTATGAACCCGGCGCGCGACTTCGCTCCTCGTCTCTTCACCTACTTCGCTGGCTACGGCTCC
AAAGTCTTCACGGAGAACGGTTACTACTTCCTGATCCCTATGTTTGCCCCGCTGATCGGCGGCGTCCT
TGGAGCGGGCATGTACGAAATCCTGGTGCAGGTGCAACATCCCCATGATCCCACAGAGTTC

Protein (315 aa)

MESSYNNTQPVIDVADKSSTYAPYADTPYSPRQGDRCERSLHHTNRSQYDRDEDHHKQWFVTRNPHLR
ECYSEFLGTFVMISFGMGVNNQVVLSEDKEGTWLSINMAWGIAVLMGVYCSEGVSGAHMNPSVTLAHA
VYGRLAWRKVPGYMLSQFLGAFVGAFAIYLLDYQRLMKADPDKETMYHNFATHPNPEINNLTAFYTEA
LATGMLLLCVYAITDQHNRSPGTVGTPFAFALMIMALGMSFGMNTGYAMNPARDFAPRLFTYFAGYGS
KVFTENGYYFLIPMFAPLIGGVLGAGMYEILVQVQHPHDPTEF

fgenesh1_pm.C_scaffold_46000013 (PrAq18)	Protein ID 71779

DNA sequence

ATGGCTCCACTGCCATCTGCAGCGCAAGCAACAGCCACGCCCGACGTGGCACCCAACGCCAGTGCCGC
GCAGTGCTTCGACGACCTGGTCGACAAGGTCGAAAGTGGCTACGCGGCCATGCCCACAACCCCCAAGA
CTGCCACCGCCGCACCGGCCTTCGCCCCCGTCAAGCCGCAGGGTCTGGCACGCTTCGCCGTGCAGAGC
GTGCACATGCGCGAGTGCTTCGCCGAGTTCCTCGGCACCTTCGTCATGATCGTCTTCGGCATGGGCGT
GAACAACCAAGTGACCAACTCGCAGGATGCCAACGGCACGTGGCTCAGCATCAACATGTGCTGGGGAA
TTGGCGTGCTCATCGGCGTGTACTGCTCCGAGGGCGTCAGCGGCGCCAACCTCAACACGGCCGTGACC
CTGGCCCACTGCGTGTACGGCCGCCTGCCGTGGTGGAAGGCGCCGGGCTACATGGTGTCGCAGCTGCT
GGGGGCCTTCTGCGGTGCCTTCGTCATCTACGTCATGCAGTACCAGAACCTGAACGTGATCGACCCGA
ACCGCGAGACGACGCAGAGCAGCTTCGCCACGTACCCTAGCGACCAGATCTCCAACTACACGGCCTTC
TACACGGAGTTCATCGGCACGGCCATGCTGGTGCTGAGCATCTACGCCATCACGGACAAGCGCAACAG
GTCTGCGGGTCCCGTGGGCTCTCCGTTTGCGTTTGCTCTCATGATCATGGCGCTGGGCATGGCGTTCG
GCATGAACACGGGATACGCTGTGAACCCTGCGCGCGACCTGGGTCCTCGCCTCTTCACGGCGGTGGCC
GGCTGGGGCTCCAAGGTCTTCACGACGCGCAACTACTACTTCTGGATCCCGCTTGTGGCCGACTCGCT
GGGCGGCGTGTGCGGCGCGGGTCTGTACCGCGTGTTGGTGGAGATCCACCACCCGCAGTTGCAGTCGC
CGCTGCTG

Protein (320 aa)

MAPLPSAAQATATPDVAPNASAAQCFDDLVDKVESGYAAMPTTPKTATAAPAFAPVKPQGLARFAVQS
VHMRECFAEFLGTFVMIVFGMGVNNQVTNSQDANGTWLSINMCWGIGVLIGVYCSEGVSGANLNTAVT
LAHCVYGRLPWWKAPGYMVSQLLGAFCGAFVIYVMQYQNLNVIDPNRETTQSSFATYPSDQISNYTAF
YTEFIGTAMLVLSIYAITDKRNRSAGPVGSPFAFALMIMALGMAFGMNTGYAVNPARDLGPRLFTAVA
GWGSKVFTTRNYYFWIPLVADSLGGVCGAGLYRVLVEIHHPQLQSPLL

fgenesh1_pg.C_scaffold_2368000002 (PrAq19)	Protein ID 87677

DNA sequence

```
ATGACTTCCAACCAAGACACAGCGGCTATTTCGTACGACGAGCTGATCGACAAGGTCGAGACGGGCTA
CGTCGACGCGCCACTGACGCCCAAGGCCGGCTCCGCGTCGCAGGGCAAGGCCGTCTCGCGCTACGCTG
TGCAGAGCGTGCACATGCGCGAGTGCCTGGCCGAGTTCCTCGGCACCTTCATCATGATCGTCTTCGGC
ATGGGCGTGAACAACCAAGTGGTGAACTCCGAGGACAAGAACGGCACGTGGCTCAGCATCAACATGTG
CTGGGGCGTGGCCGTACTCATCGGCGTGTACTGCTCCGAGGGCGTCAGTGGCGCCAACCTCAACACGG
CCGTGACCCTGGCCCACTGCGTGTACGGCCGCCTGCCGTGGTGGAAGGCGCCGGGCTACATGGTGTCG
CAGGTGCTGGGCGCCTTCTGTGGCGCCTTCATCATCTACCTGATGCAGTACCAGAACCTGAACGTGAT
CGACCCGAACCGCGAGACGATGCAGGGAAGCTGGTCCACGTACCCTAGCGACCAGATCTCCAACTACA
CGGCCTTCTACACGGAGTTCATCGGCACGGCCATGCTGGTGCTGGGCGTGTACGCCATCACGGACAAG
CGCAACAAGTCTGCCGGTCCGGTCGGCTCTCCGTTCGCCTTCTGCCTGCTCATCTGGGGACTGGGCAT
GGCGTTCGGCATGAACACGGGCTACGCCATCAACCCTGCGCGCGACTTCGGCCCGCGCCTCTTCACGT
GTCTGGCCGGCTGGGGGACCAAGGTGTTCACGCTGCGCAACCACTACTTCTGGATCCCCATCGTGGCT
CCGTTGTGCGGCGGCGTGTGCGGCGCGGGTCTGTACCGCGTGATGGTGGAGATGCACCACCCGCAGCC
GCAGCCGCGTCAG
```

Protein (299 aa)

```
MTSNQDTAAISYDELIDKVETGYVDAPLTPKAGSASQGKAVSRYAVQSVHMRECLAEFLGTFIMIVFG
MGVNNQVVNSEDKNGTWLSINMCWGVAVLIGVYCSEGVSGANLNTAVTLAHCVYGRLPWWKAPGYMVS
QVLGAFCGAFIIYLMQYQNLNVIDPNRETMQGSWSTYPSDQISNYTAFYTEFIGTAMLVLGVYAITDK
RNKSAGPVGSPFAFCLLIWGLGMAFGMNTGYAINPARDFGPRLFTCLAGWGTKVFTLRNHYFWIPIVA
PLCGGVCGAGLYRVMVEMHHPQPQPRQ
```

fgenesh1_pm.C_scaffold_660000001 (PrAq20)	Protein ID 72419

DNA sequence

```
ATGGCCATACTTCCGTCCACGCAAACACCCGATGTGGCGTCGCCCAGTCGGACCCCACAGCCTTTTGA
CGACTTGGCCGACAAGATCGAGAACGGATACACCAATATGCCGCTGACCCCCAAGGCCACGATGGGTT
TCGCCCCCGAGGGAAAACGCACAGGCCTGGCCCGCTTCGCCGTGCAGAGCGTGCACATGCGCGAGTGC
TTCGCCGAGTTCCTCGGCACCTTCGTCATGATCGTCTTCGGCATGGGCGTGAACAACCAAGTGACCAA
CTCACAGGATGCCAACGGCACGTGGCTCAGCATCAACATGTGCTGGGGAATTGGCGTGCTCATCGGCG
TGTACTGCTCCGAGGGCGTCAGTGGTGCCAACCTCAACACGGCCGTGACGCTGGCCCACTGCGTGTAC
GGCCGCCTGCCGTGGTGGAAGGCGCCGGGCTACATGATCTCGCAGCTGCTGGGGGCCTTCTGCGGTGC
CTTCGTCATCTACGTCATGCAGTACCAGAACCTGAACGTGATCGACCCGAACCGCGAGACGACGCAGA
GCAGCTTCGCCACGTACCCTAGCGACCAGATCTCCAACTACACGGCCTTCTACACGGAGTTCATCGGC
ACGGCCATGCTGGTGCTGAGCATCTACGCCATCACGGACCAGCGCAACCGTCCGGCCGGTCCCGTGGG
CGCCGCGTTCGCCTTCTGTCTCATGATCATGGCTCTGGGTATGGCGTTCGGCATGAACACGGGATACG
CCGTGAACCCTGCTCGTGACTTCGGCCCTCGCCTGTTCACGTTCTGCGCCGGCTGGGGCTCCAAGGTG
TTCACGACGCGCAACTACTACTTCTGGATCCCGCTCGTGGCCGACTCGCTGGGCGGCGTGTGCGGCGC
GGGTCTGTACCGCCTGCTCGTGGAGATCCACCACCCTCCCCTCCCGCACCAGAAC
```

Protein (313 aa)

```
MAILPSTQTPDVASPSRTPQPFDDLADKIENGYTNMPLTPKATMGFAPEGKRTGLARFAVQSVHMREC
FAEFLGTFVMIVFGMGVNNQVTNSQDANGTWLSINMCWGIGVLIGVYCSEGVSGANLNTAVTLAHCVY
GRLPWWKAPGYMISQLLGAFCGAFVIYVMQYQNLNVIDPNRETTQSSFATYPSDQISNYTAFYTEFIG
TAMLVLSIYAITDQRNRPAGPVGAAFAFCLMIMALGMAFGMNTGYAVNPARDFGPRLFTFCAGWGSKV
FTTRNYYFWIPLVADSLGGVCGAGLYRLLVEIHHPPLPHQN
```

fgenesh1_pg.C_scaffold_46000083 (PrAq21)	Protein ID 80172
DNA sequence	
ATGACCGCTCCAACCCGCAAGGTGCGCATGAGCAGCATCGAAGACGCCGCCGACCCGACGCCGTACGC CGATTTCTCCGGTTCTGCCTACGCCCCAGACGTTCGCGTTGCAATCGAGCGTGAGCCCTCACAATGCG TGATCCCCCAGAACGACTATCGCCCAACAGGTGAAGTGGCCAAAAGACGGCCCTTTGTCACGCGCAAC ACGCATTTGCGCGAGTGCTTGGCCGAGTTCCTCGCCACGACAATCGCCATCGCCTTCGGCCTGGGCGG CATCGCCCAAAGTGCCCTTTGGAACGGAGGTGGCGACACTATGACGGTCAGTGTCGGCTGGGGGGTGG CCGTCATGGTCGGCACCTTCGTGGCGGACAGCGCCAGTGGCGCCCAGATCATTTCCGTCATCACGGTC ACCAACGCCGTCTACGGCCGGATGCCGTGGTGGAAGGTGCCGGGTTACATGATAGCGCAGACATTGGG CGCCTTCGTCGGCGCGGCGCTCGTCTACGTGCTGAACGCCCAAAAGATCCGACGGGAGGACCCGAAGC AGGAAACCATGTACAAACTCTTCATCACGTACGCGCGCGACGGCGTGTCCAACTACACGGCCTTCTAC ACGGAGGTGCTGGCTTGTGCGTGCATCATGCTGGCGTCGTACGCTATCAAGGATAGGCGCAACCGCTG GCCTGGGAACAGAGGCACGCCGTTTGCGTTGGCACTGCTGGTCACGGCCATCTCGTGCGCCTTCAGCA CGAATTCGGGACTGGGCATGAGTCCGAACCGAGACTTTGGCCCTCGCCTCTTCACCTACTGCGTTGGC TACCACATGGTGTTCACGGAAGACTCGTATTACTTCTGGATCCCCATTGTGGCGCCGCTCGTCGGAGG AATCATCGGCGGCGGTCTCTACATTCTCTTCGTCGAAATGCAGCATCCAGAGCCCCCGCAGAAACCTA CCACAGAACGAGAGAGGCTCATCGAAAGCGCG	
Protein (328 aa)	
MTAPTRKVRMSSIEDAADPTPYADFSGSAYAPDVRVAIEREPSQCVIPQNDYRPTGEVAKRRPFVTRN THLRECLAEFLATTIAIAFGLGGIAQSALWNGGGDTMTVSVGWGVAVMVGTFVADSASGAQIISVITV TNAVYGRMPWWKVPGYMIAQTLGAFVGAALVYVLNAQKIRREDPKQETMYKLFITYARDGVSNYTAFY TEVLACACIMLASYAIKDRRNRWPGNRGTPFALALLVTAISCAFSTNSGLGMSPNRDFGPRLFTYCVG YHMVFTEDSYYFWIPIVAPLVGGIIGGGLYILFVEMQHPEPPQKPTTERERLIESA	

gwEuk.29.133.1 (PrAq22)	Protein ID 46593
DNA sequence	
ATGCACGTGCAATTCGACACAGAGCCCAGCACCGGACAAAGGGGCGCCTACATCGCCATCCCCCACAC GCCGTACACCTGGCGAGGATTCATCGACCGCTGCGAGCGCTCACTACACCGCACCAACCGAAGTCAGT ACGCTCGAGACGAAAACCACCACGAGAAGTGGTTCGTCACGAACCGAGCGCACGTCCGCGAGTGCCTC GCGGAGTTTCTAGGGACGTTTGTCATGATCTGCTTCGGCATGGGCGTGAACAACCAAGTCGAGCTGTC CGACGACAGCAACGGCACGTGGCTCAGCATCAACATCGCGTGGGGGATTGGCGTCCTCATGGGCGTCT ACTGCTCGGAAGGAGTGAGTGGTGCGCATCTAAACTCCGCTGTCACGCTGGCGCACGCCGTGTACGGC CGTCTGCCGTGGTGGAAGCTGCCAGGCTATTGGGTCTCGCAATTAGCGGGCGCGTTCGTGGGCGCTGC CGCCATCTACCTGCTCAACTACCAGGAGATCCTCAAGGTTGATCCGAACAAGGAAACTACGCAGTCCA ACTTTGCCACGTACCCGAGCGCCGACATCAGCAACATCACGGCCTTTTACACCGAAGCTTTGGCTGCA GGAATGCTCCTGTTGTGCATCTACGCCATCACGGACCAGAACAACCGCTCGCCCGGTACAATTGGCAC GCCGTTTGCATTTGCGCTCATGATTATGGCTCTCGGCATGAGTTTCGGCATGAACACGGGCTACGCGA TGAACCCCGCGCGTGACTTTGGCCCTCGATTGTTAACGTACGCGGTGGGCTATGGGTCGAAGGTGTGG ACGGCTGACAGCTACTACTTCTGGATTCCTATCTGCGGTCCTCTCGTGGGCGGAGTGATTGGTGCTGG CATCTACAGGCTCCTGGTCCAGGTGCAGCATCCGCACGAACCAACACTGTGTACTGTTAGT	
Protein (315 aa)	
MHVQFDTEPSTGQRGAYIAIPHTPYTWRGFIDRCERSLHRTNRSQYARDENHHEKWFVTNRAHVRECL AEFLGTFVMICFGMGVNNQVELSDDSNGTWLSINIAWGIGVLMGVYCSEGVSGAHLNSAVTLAHAVYG RLPWWKLPGYWVSQLAGAFVGAAAIYLLNYQEILKVDPNKETTQSNFATYPSADISNITAFYTEALAA GMLLLCIYAITDQNNRSPGTIGTPFAFALMIMALGMSFGMNTGYAMNPARDFGPRLLTYAVGYGSKVW TADSYYFWIPICGPLVGGVIGAGIYRLLVQVQHPHEPTLCTVS	

fgenesh1_pm.C_scaffold_6000003 (PrAq23)	Protein ID 71169

DNA sequence

ATGGCAAGGAACTCGCACCTTGAGAAGAGCGACCACGGGTACCTGGAAGCGGGCACCCCGGAGCAGGG
CGACCTGCAGTACCTTACGCCCGTGCACAACGTACCGTACATGGTCAAGAGCCAATTCGCCAAGGAGA
TGATGGCCGAGTTCGTCGCCACATTCGTGACCATGCTCTTCGGGCTGTCGTGCATGACGCAGACCGTG
CTGAGTAGCGAGGCCAGTGGCAGCTTCGTGACTATCGCGCTGTGCTGGGGACTCGCCTTCTTCTTCGG
CATCACCATTGGTGGCGGCGTGTCCGGTGCACACCTGAACCCGGCCGTGACCACGTCGCTGGCGCTGC
TCAAGCTGCTGCCGTGGAAGAAGGTGCCGTTCTACATCCTCAACCAGGTCGTGGCCGCCTTTGTGGCT
GCGCTCTTCGTGTACATCTTGTACCGGCCAATGTTTAATGAGGTGGACCCGGACCGGACGTCCACGCA
CACCATCTTCGCCACGTTCCCGCACGAGAACGTGGGCAACTTCACGTGCTTTTTGACCGAGTTCGTGG
CCACGGCGCTCCTGATCCTGGGCATCCTGGCGCTGCTGGACCAGCACAACCGCCCGATCGGCCCCAAG
GCCGCGCCCGCCGCCGTTGGAGCGTTGGTGAGCACCATCGCCATGGGTTTCGCCATGAACACGGGGCT
GGCCATCAACCCCGCGCGCGACCTGGGGCCTCGTCTCTTCATGCTGTGCGCTGGCTGGGGCTCGCGCG
TGTTCTCGCTCAACCATTACTACTTCTGGGTGCCGATTGTGGCTCCGATCACGGGAGGCGCTACGGGC
GCCTTCGTGTACGAAGCGATGATCAGCTACCACCACCCGGAGAAGAACCGTGGGCCGCACCCGGAGTA
CCGATTC

Protein (297)

MARNSHLEKSDHGYLEAGTPEQGDLQYLTPVHNVPYMVKSQFAKEMMAEFVATFVTMLFGLSCMTQTV
LSSEASGSFVTIALCWGLAFFFGITIGGGVSGAHLNPAVTTSLALLKLLPWKKVPFYILNQVVAAFVA
ALFVYILYRPMFNEVDPDRTSTHTIFATFPHENVGNFTCFLTEFVATALLILGILALLDQHNRPIGPK
AAPAAVGALVSTIAMGFAMNTGLAINPARDLGPRLFMLCAGWGSRVFSLNHYYFWVPIVAPITGGATG
AFVYEAMISYHHPEKNRGPHPEYRF

fgenesh1_pg.C_scaffold_114000014 (PrAq24)	Protein ID 84475

DNA sequence

ATGGAGCGCGGAGATCCCACGCCCGCCCTTCGAGCTCGCCGAGGTCACTTTCCAATCCGCCTCCATTC
GCTCACGATGGCTCGCTTATCCGACAGCGGCAACCCCAAGGACAAAGACGGCTACGTGGGGCTCCAAG
AGGCCGACGTGGCCATGACGCCTGGCCCCGCCAAACCGTTCCAAGTCAAGTCACCGCTGCTGCGCGAG
TGCATGGCCGAGTTCATCGGCACCATGGTGCTCATTCTGTTCGGCGACGGCGTCGTGGCCCAAGTCGT
GCTCGGCGAGGGCACCAAGGGCGAGTACATCAACATCAACTTGTGCTGGGGTCTCGGTGTGCTCTTCG
GCATCCACGCGTCGGGCGGCGTCTCCGGTGCACACCTGAACCCTGCGGTGACAACAACCCTCGCGATC
TACGGACGACTGGAATGGCGCAAAGTGATCCCGTATATCGCCGCCCAAGTGTTGGGCGCCTTTGTGGC
GGCCTTCATCGTCTGGGCCGTCTACTGCCCCATGTTCAATGCTATCGACCCTGATAAGGAATCCACGC
AGGGCGTCTTCGCCACGTACCCGTACAGCAGCGACGTGCCGGTGGGGACGTGTTTTCTGACGGAAGTG
GTGGGCACGGCCTTGCTGCTCGGAGGCATCTTCGCCATCGGAGACGAGCTCAACAAGCCCGCCAGTCC
GTACTCGCAGCCGAGCGCGGTGGCGTTGCTGGTCGTGGCCATTGGCATGGCGTTCGGCATGAACTCCG
GCTACGCTATCAACCCTGCGCGTGACTTCGGGCCGCGCCTGTTCTCGCTCTGCGCTGGCTGGGGGTCC
AAGGTCTTCACCATGCGCGACCACTACTTCTGGGTGCCGATCATCGGCCCGCTGCTCGGCGGTGCCAT
CGGTGGCGGCGTGTACATTGGACTCGTGGAGCACCACCACCCGCGCGACTACAAGCACCCGCTGGAGA
AC

Protein (318 aa)

MERGDPTPALRARRGHFPIRLHSLTMARLSDSGNPKDKDGYVGLQEADVAMTPGPAKPFQVKSPLLRE
CMAEFIGTMVLILFGDGVVAQVVLGEGTKGEYININLCWGLGVLFGIHASGGVSGAHLNPAVTTTLAI
YGRLEWRKVIPYIAAQVLGAFVAAFIVWAVYCPMFNAIDPDKESTQGVFATYPYSSDVPVGTCFLTEV
VGTALLLGGIFAIGDELNKPASPYSQPSAVALLVVAIGMAFGMNSGYAINPARDFGPRLFSLCAGWGS
KVFTMRDHYFWVPIIGPLLGGAIGGGVYIGLVEHHHPRDYKHPLEN

fgenesh1_pm.C_scaffold_114000006 (PrAq25)	Protein ID 72237
DNA sequence	
ATGTCGGAGTACGGCTCACTTAATAAGCCTGGATACGCTGATCTCGAGGGCAACCCGCGTCTGCCGTT CTACGCCGTGCGCAACCCGGACGTCCGCGCCTACCTGGCCGAATTCGTCGGCACTTTCATCCTCGTCC TCATCGGTGACGGCTCCGTGGCCCAGTACGTGCTGGGCGGTGGCAACGCTGGCCACTACCTGTCGGTG AACCTTGCCTGGGGTATCGCTTTGCTGTTCGGCATTCACTTCTCGGGCGGTGTCAGCGGTGGCCACCT CAACCCTGCAGTAACGCTGACCCTCGCTGCATTTGGGCGCTTCGAGTGGTACAAGCTTCCGGGCTACT TCGTGGCCCAGACCCTCGGGGCCTTCGCGGCTGCCTGGGTAGTGTTCGTTGTCTACTACCCCTGGTTC GACGTTCAGGACCCGGAGCGTGAGACCACCCAGGGCATCTTCGCCACATACCCGAACGAGCAGATCCC CAACTGGTCCGGACTCGCTAACGAGATCGTCGGCACAGCTCTGCTGGTTAGCGGCATCTTCGCTGTCG GCGACCAACTCAACAAACCTGCCAGTCCCTACAGCTTCCCCGGAGCTGTTGCTCTCATGCTCACGTGC ATTGGCATGGCGTTCGGCCTGGACACGGGCTACGCGCTCAACCCCGCCCGTGACTTCGGTCCTCGCCT CTTTACGTTCTTCGCTGGCTGGGGCTGGAAGGTGTTCGCAGCTCGCAGCTTTTACTTCTGGATCCCTA TCGTCGGTCCTTTCATCGGCGGTCTGCTTGGTGCTGGTCTGTACGTCGGTCTCATCGAGAACTTCCAC CCCCGCGAG	
Protein (275 aa)	
MSEYGSLNKPGYADLEGNPRLPFYAVRNPDVRAYLAEFVGTFILVLIGDGSVAQYVLGGGNAGHYLSV NLAWGIALLFGIHFSGGVSGGHLNPAVTLTLAAFGRFEWYKLPGYFVAQTLGAFAAAWVVFVVYYPWF DVQDPERETTQGIFATYPNEQIPNWSGLANEIVGTALLVSGIFAVGDQLNKPASPYSFPGAVALMLTC IGMAFGLDTGYALNPARDFGPRLFTFFAGWGWKVFAARSFYFWIPIVGPFIGGLLGAGLYVGLIENFH PRE	

fgenesh1_pm.C_scaffold_114000005 (PrAq26)	Protein ID 72236
DNA sequence	
ATGACCGAGTACGGCACCAGCAAGCACGGCTACGCCGACCTGGAGGGCAACCAGCGTCTGCCGTTCTA CCACGTCCGCAACCCGGACTTCCGCGCCTACCTGGCCGAATTCGTCGGCACTTTCATCCTCGTACTCA TCGGTGATGGTTCTGTGGCGCAGTACGTGCTGGGCGGTGGTGGTGCCGGCCACTACCTGTCGGTGAAC CTGTGCTGGGGCATCGCTCTGCTGTTCGGCATCCACTTCTCCGGAGGTGTCAGTGGTGGGCACCTCAA CCCCGCCGTCACCCTCACGCTCGCCCTGTTCAAGCGTTTCGAGTGGCACAAGGTGCCCGGCTACTTCG TGGCCCAGACCGCCGGAGCTTTCGCGGCCGCCATGATCGTTTTCATCGTGTACTACCCGTGGTTCGAT ATCCACGACCCGGAGCGTGAGACCACCCAGGGCATCTTCGCCACATACCCGAACGAGCAGATCCCTAA TTGGTCAGCGCTTGCAAACGAGATCATCGGCACGGCTCTGCTGGTTAGCGGCATCTTCGCTGTCGGCG ACCAACTCAACAAACCTGCCAGTCCCTACAGCTTCCCCGGAGCCGTTGCTCTCATGCTCACGTGCATT GGCATGGCGTTCGGCCTGGACACGGGCTACGCGCTCAACCCCGCCCGTGACTTCGGTCCTCGCCTCTT TACGTTCTTCGCTGGCTGGGGCTGGAAGGTGTGGACTCTGCGTGACGGCTACTTCTGGATCCCCATCG TTGGCCCGTTTGTCGGCGGAATTCTTGGTGCTGCCACCTACGTCGGCCTCATTGAGGCCCACCACCAA CCTCAG	
Protein (274 aa)	
MTEYGTSKHGYADLEGNQRLPFYHVRNPDFRAYLAEFVGTFILVLIGDGSVAQYVLGGGGAGHYLSVN LCWGIALLFGIHFSGGVSGGHLNPAVTLTLALFKRFEWHKVPGYFVAQTAGAFAAAMIVFIVYYPWFD IHDPERETTQGIFATYPNEQIPNWSALANEIIGTALLVSGIFAVGDQLNKPASPYSFPGAVALMLTCI GMAFGLDTGYALNPARDFGPRLFTFFAGWGWKVWTLRDGYFWIPIVGPFVGGILGAATYVGLIEAHHQ PQ	

G.7 VMD DB analysis – *P. infestans*

CL1Contig5332 (PiAq8)

DNA sequence

```
ATGGCTCCGCTCCTCTCCAACCAGACCCCGGATGAGTCAGCTAACTATGCCCCGCAGCCTTTCGAAGA
CATCGACGGCAAGCACGAGGTCAGCTACATCGATATGCCATCGACTCCTAAGGCTGACACCAGCCTCG
CTCGCTTTGCCGTGCAGAGTGTGCACTTACGCGAGTGTTTCGCCGAGTTCCTCGGCACCTTCGTCATG
ATCTTGTTCGGTATGGGCGTCAACAACCAGGTGACCAACTCACAAGACGCCAACGGCACGTGGCTCAG
TATCAACATGTGCTGGGGTATCGGAGTCTTAATTGGTGTGTACTGCACGGAAGGAATCAGCGGAGCCC
AGTTGAACACTGCAGTGACACTGGCACACTGTGTATACGGCCGTCTACCTTGGTGGAAGGCGCCAGGC
TACATGATCTCACAGGTCTTGGGTGCATTCGTGGGCGCATTTGTGATCTGGGTCATGCAGTGGCAGAA
CCTGAACGCCATTGACCCCGACCGAGAGACTACCCAGAGCAGCTTCGCTACTTACCCGAGTGACAACA
TTTCTAATTACACGGCCTTCTACACGGAGGTAGTCGGCACGGGGATGTTGTTACTCGGTATCTATGCC
ATCACGGATCAAAGGAACCGTCCGGCTGGCCCTGTAGGCGTACCATTTGCCTTCTGCCTTTTGATCAT
GGCGCTCGGTATGTGCATTGGTATGAACACGGGTTACGCTATTAACTCGGCGCGTGACTTCGGCCCTC
GTCTTTTCACGAGCATCGCTGGCTGGGGATCCAAGGTATTCACCCTGCGCGATCACTACTTTTGGATC
CCGATCGTGGGTCCACTGGTTGGGGGCGTGATCGGTGCGGGTCTGTACATTACGCTATTAATAGTCGA
GATGCACCACCCGCGTCGGGCACTTCCGTAA
```

Protein (304 aa)

```
MAPLLSNQTPDESANYAPQPFEDIDGKHEVSYIDMPSTPKADTSLARFAVQSVHLRECFAEFLGTFVM
ILFGMGVNNQVTNSQDANGTWLSINMCWGIGVLIGVYCTEGISGAQLNTAVTLAHCVYGRLPWWKAPG
YMISQVLGAFVGAFVIWVMQWQNLNAIDPDRETTQSSFATYPSDNISNYTAFYTEVVGTGMLLLGIYA
ITDQRNRPAGPVGVPFAFCLLIMALGMCIGMNTGYAINSARDFGPRLFTSIAGWGSKVFTLRDHYFWI
PIVGPLVGGVIGAGLYITLLIVEMHHPRRALP
```

CL51Contig1 (PiAq18)

DNA sequence

```
ATGGCAGAATACAAGCCTGGTTTCGGTGGCTATGCTGACCTCGAAGGCAATGGTCGACTAGCACCCTT
CGCGATTCGTAACCCAGACATGCGCGCTTACATGGCTGAATTCGTCGGTACTTTTATCCTCGTGCTGA
TCGGTGACGGCTCCGTGGCCCAGTTCGTGCTTAGTAAGAGGGCTGCTGGTGACTATCTTTCGGTAAAT
CTCTGCTGGGGTATCGCACTTCTCTTCGGGGTTCATTTCTCGGGTGGTGTTAGCGGTGGTCATCTCAA
TCCTGCAGTCAGTGTGGCCATGGCAATATTCAAGCGCTTCGAGTGGCGCAAAGTTCCGGGCTACATCA
TCGCGCAGACAATGGGGGCGTTCGCTGCTGCGTTGGTGCTCTTCATCGTCTACTACCCGTGGCTCGAC
ATCGTCGACCCCGAGCGCGAATTCACTCAGGGAATCTTTGCCACATACCCAAACCCACAGATCCCGAA
CTGGGCTGCGTTCGCTAACGAAGTCATCGGTACGGCTCTGCTAGTTAGTGGCATCTTCGCGTTGTGTG
ACCAAATCAACAAACCTGCCAGTCCCTACAGTTTCCCCGGTGCTGTGGGTCTACTACTGACAGGCATC
GGTATGTCGTTCGGCCTCAATACAGGCTACGCTCTTAACCCCGCTCGCGACTTCGGTCCGCGCCTTCT
GACGTTTTTCGGCGGCTGGGGTTGGAAGGTGTTCTCGTCGCACGGTGGGTACTTCTGGATCCCCATTT
TGGGCCCGTTCATAGGTGCAGTGCTGGGTGCGGCCATGTATGTAGGCCTCGTTGAGCTCCACCACCCG
CCCCAGTAG
```

Protein (274 aa)

```
MAEYKPGFGGYADLEGNGRLAPFAIRNPDMRAYMAEFVGTFILVLIGDGSVAQFVLSKRAAGDYLSVN
LCWGIALLFGVHFSGGVSGGHLNPAVSVAMAIFKRFEWRKVPGYIIAQTMGAFAAALVLFIVYYPWLD
IVDPEREFTQGIFATYPNPQIPNWAAFANEVIGTALLVSGIFALCDQINKPASPYSFPGAVGLLLTGI
GMSFGLNTGYALNPARDFGPRLLTFFGGWGWKVFSSHGGYFWIPILGPFIGAVLGAAMYVGLVELHHP
PQ
```

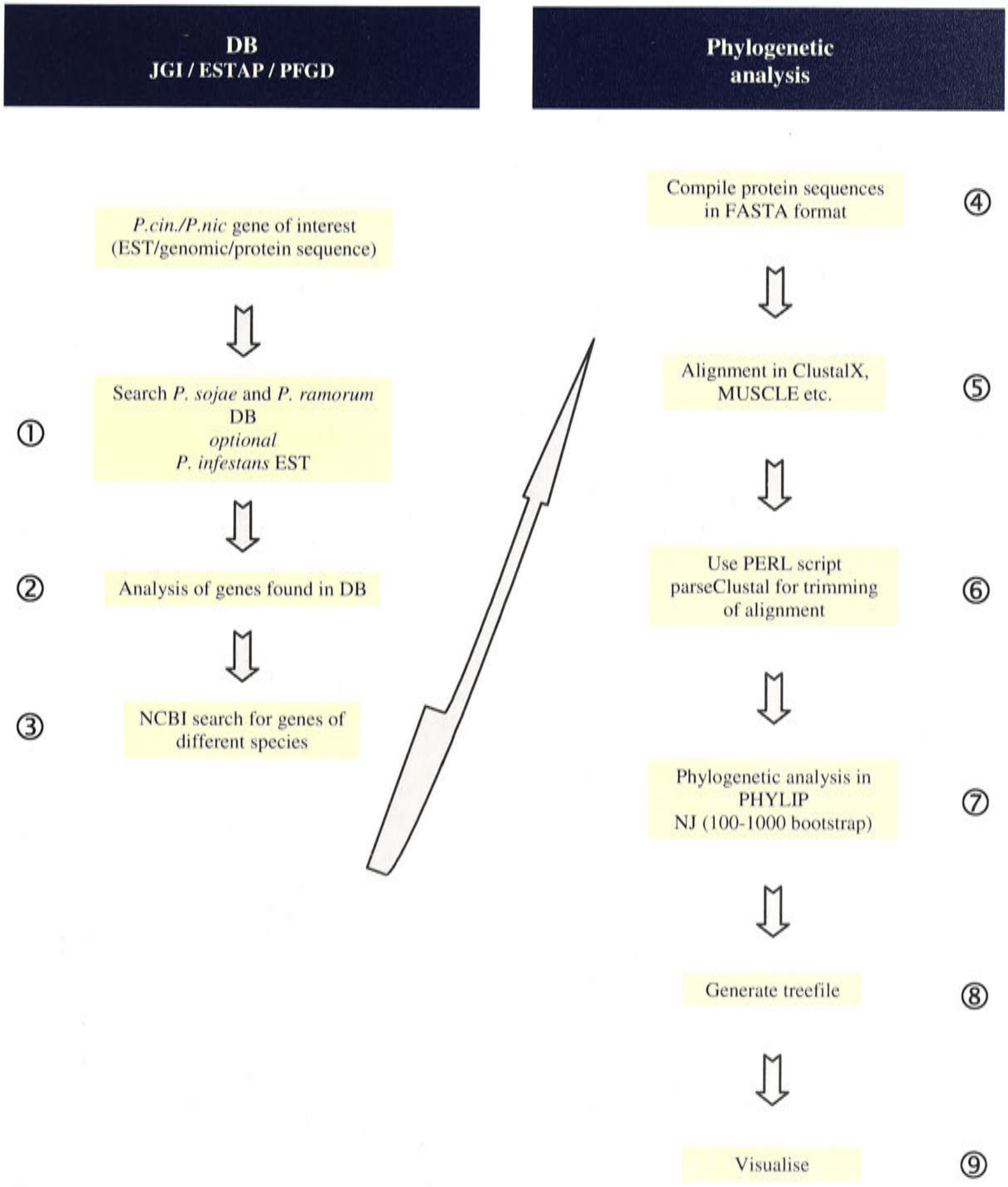
G.8 NCBI sequences

Arabidopsis thaliana AQPb	CAB51216
Arabidopsis thaliana MIPa	CAB10515
Arabidopsis thaliana MIPb	AAC62778
Arabidopsis thaliana TIP	AAC42249
Arabidopsis thaliana AQPa	CAB41102
Arabidopsis thaliana NOD26a	CAB39791
Arabidopsis thaliana NOD26b	CAA16760
Arabidopsis thaliana NOD26c	CAA16748
Arabidopsis thaliana PIP	AAB65787
Arabidopsis thaliana PIP1.2	Q06611
Arabidopsis thaliana PIP1.3	Q08733
Arabidopsis thaliana αTIP	P26587
Arabidopsis thaliana βTIP	AAB84183
Arabidopsis thaliana γ2TIP	AAC62397
Arabidopsis thaliana γTIP	P25818
Arabidopsis thaliana δTIP	AAC49992
Arabidopsis thaliana SIP1.1	Q9M8W5
Arabidopsis thaliana SIP1.2	Q9FK43
Arabidopsis thaliana SIP2.1	Q9M1K3
Aspergillus fumigatus	XP_750737
Beta vulgaris PIP1	AAB67868
Beta vulgaris PIP2	AAB67869
Beta vulgaris PIP3	AAB67870
Brucella abortus	AAF73105
Caenorhabditis elegans AQP9	NP_001021552
Caenorhabditis elegans AQP10	NP_496105
Caenorhabditis elegans AQP11	NP_499821
Candida albicans	XP_715780
Chlamydomonas reinhardtii	AAP33478
Danio rerio	CAE50608
Dictyostelium discoideum	BAA85158
Dictyostelium discoideum	AAB72014
Encephalitozoon cuniculi	NP_586002
Escherichia coli AQPZ	NP_415396
Escherichia coli GlpF	NP_418362
Homo sapiens AQP0	P30301
Homo sapiens AQP1	P29972
Homo sapiens AQP2	AAD38692
Homo sapiens AQP3	AAY68214
Homo sapiens AQP4	AAB26958
Homo sapiens AQP5	NP_001642
Homo sapiens AQP6	Q13520
Homo sapiens AQP7	O14520
Homo sapiens AQP8	NP_001160

Homo sapiens AQP9	O43315
Homo sapiens AQP10	NP_536354
Homo sapiens AQP11	NP_766627
Homo sapiens AQP12	NP_945349
Hordeum vulgare MIPa	Q40047
Hordeum vulgare MIPb	CAA54233
Leishmania major	AAS73184
Magnaporthe grisea	EAA50403
Mus musculus AQP1	Q02013
Mus musculus AQP3	AAH27400
Mus musculus AQP7	BAA24537
Mus musculus AQP9	NP_071309
Neurospora crassa	EAA35947
Nicotiana tabacum AQP1a	CAA04750
Nicotiana tabacum AQP1b	CAA69353
Nicotiana tabacum AQPGLY	CAB40742
Nicotiana tabacum TIP	P21653
Oryza sativa MIPa	AAB18817
Oryza sativa MIPb	Q40746
Oryza sativa PIP1a	CAA11896
Oryza sativa PIP2a	AAC16545
Oryza sativa TIP	P50156
Plasmodium falciparum	CAC88373
Plasmodium yoelii	PY05950
Rattus norvegicus AQP1	P29975
Rattus norvegicus AQP3	NP_113891
Rattus norvegicus AQP7	P56403
Rattus norvegicus AQP9	AAC36020
Saccharomyces cerevisiae AQY1	NP_015518
Saccharomyces cerevisiae AQY2	AAD10058
Saccharomyces cerevisiae FPS1	CAA38096
Saccharomyces cerevisiae YFL054	BAA09187
Toxoplasma gondii	CAE46485
Trypanosoma brucei 1	CAG27020
Trypanosoma brucei 2	CAG27021
Trypanosoma brucei 3	CAG27022
Trypanosoma cruzi 1	XP_813167
Trypanosoma cruzi 2	XP_816487
Trypanosoma cruzi 3	AAM76680
Xenopus laevis	CAA10517

Appendix H Database and phylogeny manual

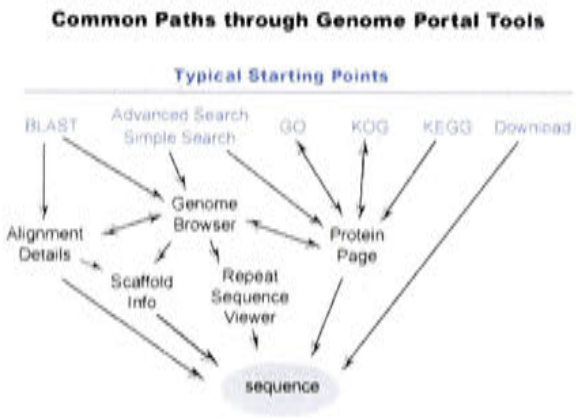
Flow-Chart



①

- Go to respective DB
- Use 'Advanced Search' for keyword search or 'BLAST' for blasting of nucleotide or protein sequence

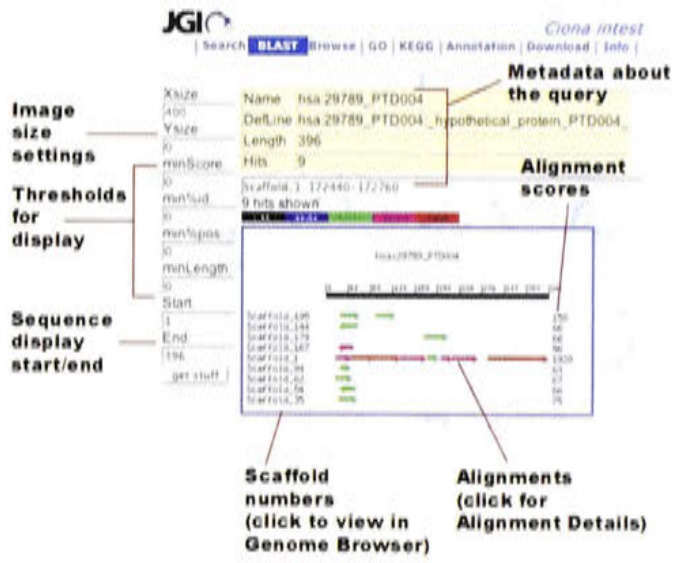
- For JGI help file see http://genome.jgi-psf.org/help/



②

- Create Excel sheet
- Output for keyword search:

- Output for BLAST search:



(for protein page go to scaffold → click on gene model)

- Note details of genome and protein pages (A - ①②, C - ⑤⑥⑦)
- Save protein page and genome browser as PDF files for future reference (optional)
- Check transcript prediction (A – 3 → B)
  - Pay special attention to predicted introns [rec. seq. usually GTRYG …. YAG]

    Y – pyrimidine (T / C)

    R – purine (A / G)

- If in doubt about predicted ATG check for available EST sequences. also consider codon usage → GC content higher in coding regions than in non-coding regions
- Use the "Replace" function in MSWord for easy formatting of sequences

  ^p – paragraph mark

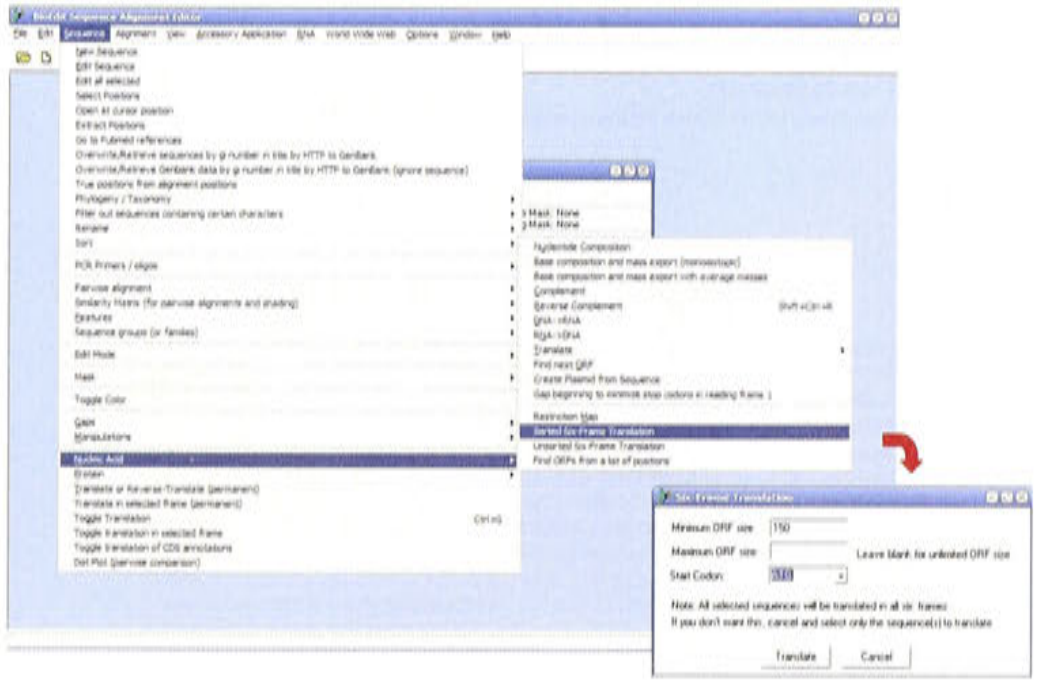
  ^w – white space

  wildcards:

  ? any single character

  [ ] one of the specified characters
- Increase upstream/downstream padding if in doubt of correct transcript (B - ④)
- Use programs like Bioedit / DNAstar etc. for easy 6-frame translation

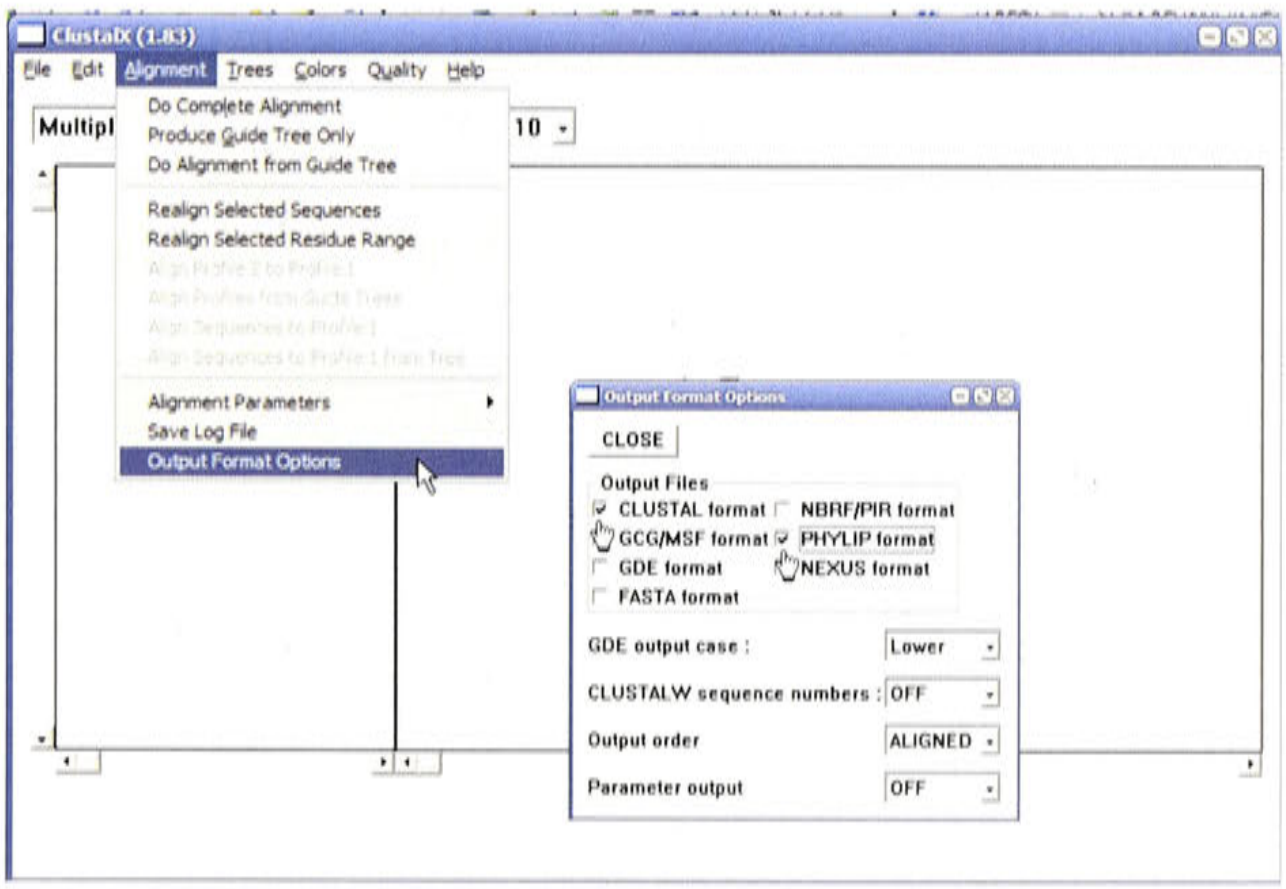
  or go to http://molbiol.ru/eng/scripts/01_13.html



- Include final protein sequence in Excel sheet
- Check for duplicates by using a multiple sequence alignment program like ClustalX.

  if in doubt about a gene (e.g. real recent gene duplication or assembly error – remember the whole assembly was done by a machine!!!) Check identity of 3'/5' UTRs, compare to EST sequences if available etc.

P.infestans EST DB ("Phytophthora DB" → "Toolkit" → select Database **OR** "PFGD")

	• BLAST or Keyword search • Only close to full length gene contigs are useful for later analysis • Note details as above
③	• Obtain additional gene related sequences of other species by blasting NCBI • Save protein sequences in FASTA format
④	• Compile all protein sequences in a single file in FASTA format (note: ClustalX does not accept spaces, use _ instead) • Save as .txt
⑤	• Align sequences with ClustalX, MUSCLE or similar • For ClustalX, use default settings but save file in Clustal (.aln) AND PHYLIP (.phy) format 
⑥	• Use Perl-script "parseClustal" (courtesy of Allan Dickerman, VBI, Blacksburg, VA, USA) to trim off gaps • Install Perl • *Windows:* Run 'cmd', change to location of "parseClustal.pl" (① - To change drive and directory at the same time, use cd /D <driveletter>:\<folder> *Mac:* Run the "Terminal" application, change to location of "parseClustal.pl" ② - Type "perl parseClustal.pl" • Move the .phy file from the Clustal alignment into the same folder as "parseClustal.pl"

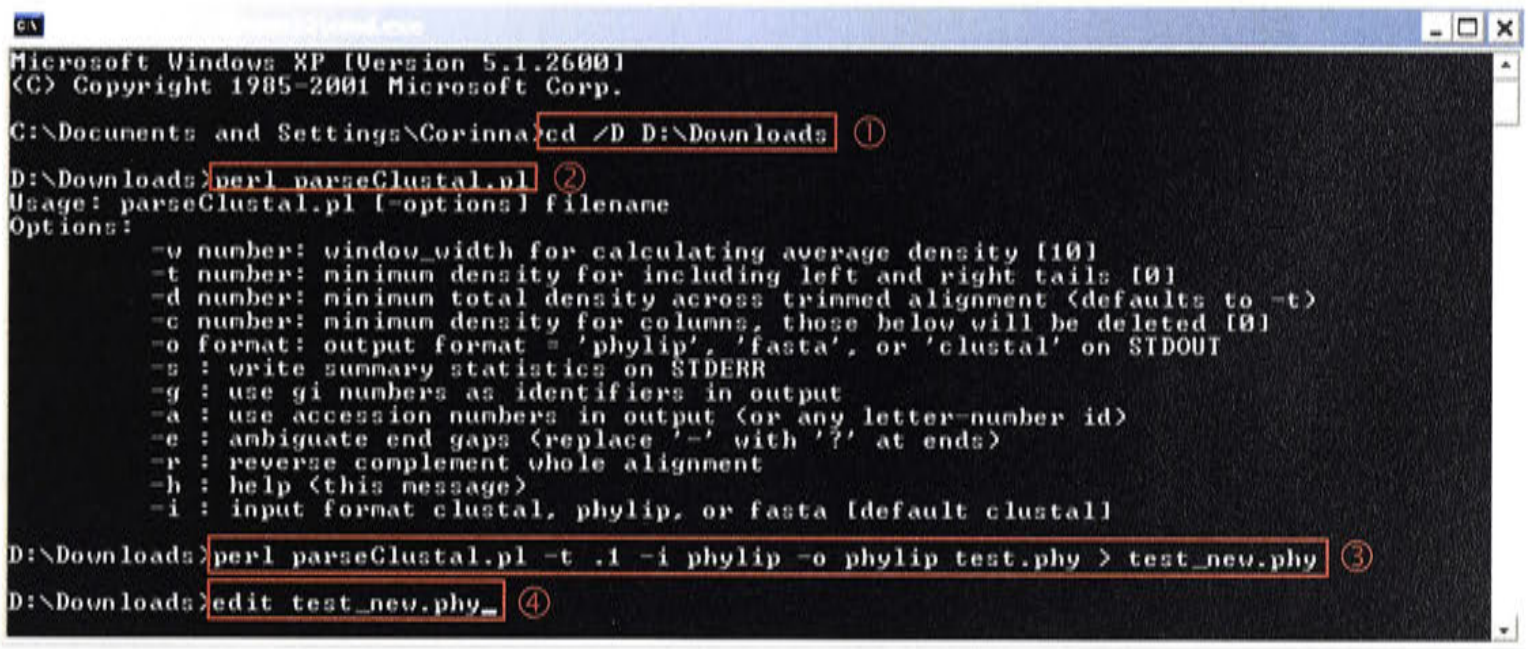
- ③ - Use the following command line to run the program:

  perl parseClustal.pl –t .1 –i phylip–o phylip 'input file' > 'output file'

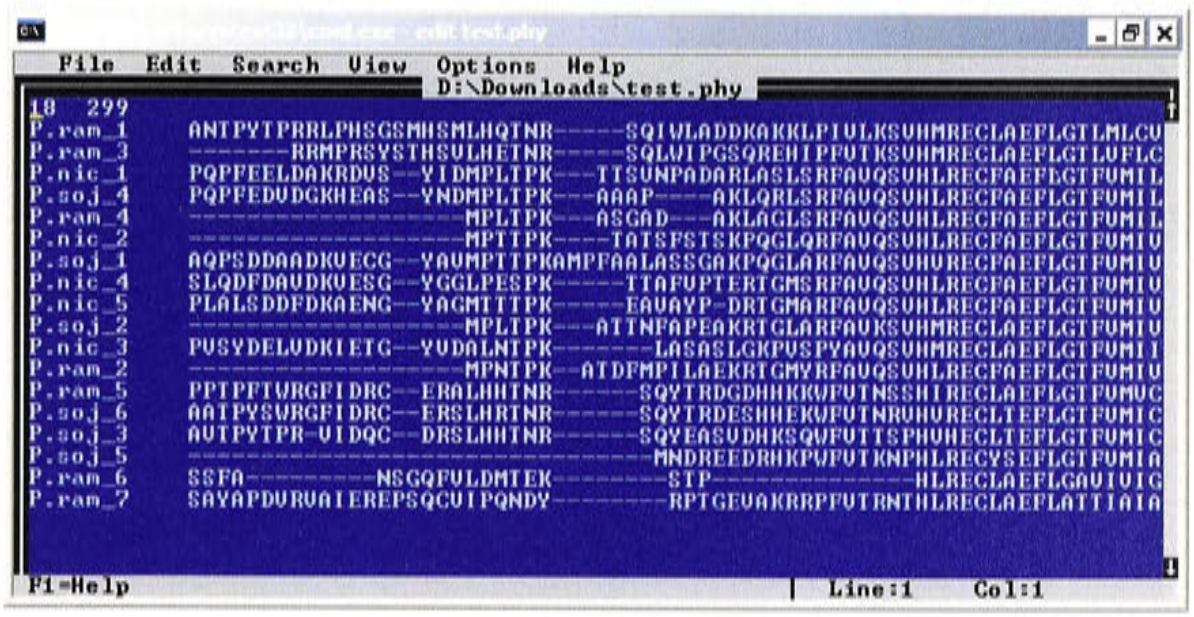
  Options:

  -t : removes C/N terminal tails

  -c : eliminates in-sequence gaps



- *Note:* If you are using a different program for alignment make sure that the output file is in clustal, phylip or fasta format. Adjust perl command line accordingly for example change to fasta input file "-i fasta"
- ④ - Type "edit <filename>" to view file content



⑦ + ⑧	• Use PHYLIP or similar phylogenetic analysis program (Batch files for automating the analysis are available upon request) (Documentation: http://evolution.genetics.washington.edu/phylip/phylip.html) • Instructions below are for PHYLIP

SEQBOOT (produces multiple data sets from original data set by bootstrap resampling)

- Rename the ClustalX .phy file as "infile"
- Move it to the PHYLIP folder which contains the program "SEQBOOT"
- Start "SEQBOOT"; use the following settings:

```
Bootstrapping algorithm, version 3.64

Settings for this run:
  D       Sequence, Morph, Rest., Gene Freqs?  Molecular sequences
  J  Bootstrap, Jackknife, Permute, Rewrite?   Bootstrap
  %    Regular or altered sampling fraction?   regular
  B      Block size for block-bootstrapping?   1 (regular bootstrap)
  R                    How many replicates?    1000
  W                Read weights of characters? No
  C                  Read categories of sites? No
  S     Write out data sets or just weights?   Data sets
  I              Input sequences interleaved?  Yes
  0       Terminal type (IBM PC, ANSI, none)?  IBM PC
  1        Print out the data at start of run  No
  2     Print indications of progress of run   Yes

  Y to accept these or type the letter for one to change
```

- Rename "outfile" to "infile"

PROTDIST (construction of a distance matrix)

- Start "PROTDIST"; use the following settings:

```
Protein distance algorithm, version 3.65

Settings for this run:
  P  Use JTT, PMB, PAM, Kimura, categories model?  Dayhoff PAM matrix
  G  Gamma distribution of rates among positions?  No
  C            One category of substitution rates? Yes
  W                     Use weights for positions? No
  M                   Analyze multiple data sets?  Yes, 1000 data sets
  I                 Input sequences interleaved?   Yes
  0               Terminal type (IBM PC, ANSI)?    IBM PC
  1          Print out the data at start of run    No
  2        Print indications of progress of run    Yes

Are these settings correct? (type Y or the letter for one to change)
```

- Rename "outfile" to "infile"

NEIGHBOR (NEIGHBOR constructs a tree by successive clustering of lineages, setting branch lengths as the lineages join)

- Start "NEIGHBOR"; use the following settings:

```
Neighbor-Joining/UPGMA method version 3.64

Settings for this run:
  N       Neighbor-joining or UPGMA tree?  Neighbor-joining
  O                       Outgroup root?   No, use as outgroup species  1
  L          Lower-triangular data matrix? No
  R          Upper-triangular data matrix? No
  S                       Subreplicates?   No
  J     Randomize input order of species?  Yes (random number seed =     503)
  M           Analyze multiple data sets?  Yes, 1000 sets
  0   Terminal type (IBM PC, ANSI, none)?  IBM PC
  1    Print out the data at start of run  No
  2  Print indications of progress of run  Yes
  3                       Print out tree   Yes
  4        Write out trees onto tree file? Yes


  Y to accept these or type the letter for one to change
```

- Rename "outtree" to "intree"

CONSENSE (reads a file of computer-readable trees and constructs a consensus tree)

- Move the file "intree" from above to the PHYLIP folder which contains the program "CONSENSE"
- Start "CONSENSE"; use the following settings:

```
Consensus tree program, version 3.64

Settings for this run:
 C         Consensus type (MRe, strict, MR, Ml):  Majority rule (extended)
 O                              Outgroup root:  No, use as outgroup species  1

 R                 Trees to be treated as Rooted:  No
 T           Terminal type (IBM PC, ANSI, none):  IBM PC
 1               Print out the sets of species:  Yes
 2         Print indications of progress of run:  Yes
 3                             Print out tree:  Yes
 4                Write out trees onto tree file:  Yes

Are these settings correct? (type Y or the letter for one to change)
```

⑨

- Visualise tree with tree viewing software (e.g TreeView, MEGA)

Useful paper and manuals

JGI Helpfile	http://genome.jgi-psf.org/help/
PHYLIP manual	http://evolution.genetics.washington.edu/phylip/phylipweb.html
"Phylogeny for the faint of heart: a tutorial"	Baldauf, SL Trends in Genetics Vol.19 (6): 345-351 (2003)

Programs

MUSCLE	http://www.drive5.com/muscle/	Phylogeny
MrBayes	http://mrbayes.csit.fsu.edu/index.php	Phylogeny
PHYLIP	http://evolution.genetics.washington.edu/phylip.html	Phylogeny
HyPhy	http://www.hyphy.org/	Phylogeny
ClustalX	http://bips.u-strasbg.fr/fr/Documentation/ClustalX/	Seq. alignment
BioEdit	http://www.mbio.ncsu.edu/BioEdit/bioedit.html	Seq. alignment etc
pDraw32	http://www.acaclone.com/	Seq. alignment etc
Phylodendron	http://iubio.bio.indiana.edu/soft/molbio/java/apps/trees/	Tree view
TreeView	http://taxonomy.zoology.gla.ac.uk/rod/treeview.html	Tree view